

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2005, 20:55:08 ; Search time 52 Seconds

(without alignments)
2697.415 Million cell updates/sec

Title: US-10-607-631-8

Perfect score: 9732

Sequence: 1 MNKKSTLLATATAAIIIGST.....TNNAFNNVFKEFNISKIIVE 1879

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTCUS-COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	301.5	3.1	2662	4	US-09-595-684B-31
2	301	3.1	2663	4	US-09-538-092-1252, Ap
3	286.5	2.9	1979	4	US-09-949-016-6468
4	284.5	2.9	2047	4	US-09-949-016-7404
5	276	2.8	3259	4	US-09-949-016-6507
6	275.5	2.8	1663	5	PCT-US93-07261-16
7	275.5	2.8	10182	3	US-09-134-001C-3159
8	271	2.8	2733	4	US-09-949-016-11433
9	262.5	2.7	5024	4	US-09-710-279-2964
10	256.5	2.6	1588	5	PCT-US93-07261-11
11	256.5	2.6	3696	3	US-09-134-001C-5080
12	256	2.6	3418	3	US-08-755-587-44
13	255	2.6	2482	1	US-08-328-254-6
14	253	2.6	3418	2	US-08-639-501-2
15	253	2.6	3418	3	US-09-044-946-2
16	253	2.6	3418	3	US-09-044-908-2
17	252	2.6	2285	3	US-09-308-375-2
18	252	2.6	2285	4	US-09-932-183A-2
19	250.5	2.6	2710	1	US-08-480-604A-6
20	250.5	2.6	2710	2	US-08-405-496A-6
21	250.5	2.6	2710	3	US-08-915-136-6
22	250.5	2.6	2710	4	US-08-957-310-6
23	250.5	2.6	2710	3	US-10-011-366-6
24	250.5	2.6	2710	4	US-09-084-517-6
25	250.5	2.6	3418	2	US-08-603-753D-4
26	250.5	2.6	3418	3	US-09-099-753-4
27	250.5	2.6	3418	3	US-08-986-106-4

28	245	2.5	904	3	US-09-198-484-2	Sequence 2, Appli
29	244.5	2.5	2329	3	US-08-755-587-16	Sequence 16, Appli
30	242.5	2.5	2954	4	US-09-150-867-1	Sequence 1, Appli
31	238.5	2.5	3913	4	US-09-949-016-10933	Sequence 10933, A
32	238.5	2.5	4377	4	US-09-949-016-6978	Sequence 6978, Ap
33	233	2.4	2807	4	US-09-543-681A-4980	Sequence 4980, Ap
34	231	2.4	3210	4	US-09-538-092-1154	Sequence 1154, Ap
35	231	2.4	3878	4	US-09-914-269-11	Sequence 11, Appli
36	229.5	2.4	2472	4	US-09-538-092-1312	Sequence 1312, Ap
37	228	2.3	3248	1	US-08-353-700-1	Sequence 1, Appli
38	228	2.3	3248	5	PCT-US95-16216-1	Sequence 1, Appli
39	227.5	2.3	1010	3	US-09-134-001C-5178	Sequence 5178, Ap
40	227	2.3	2777	4	US-09-543-681A-6124	Sequence 6124, Ap
41	226.5	2.3	3135	1	US-08-323-170B-2	Sequence 2, Appli
42	226.5	2.3	3135	3	US-08-954-441-2	Sequence 2, Appli
43	226	2.3	1935	4	US-09-538-092-916	Sequence 916, App
44	225.5	2.3	2186	4	US-09-949-016-10828	Sequence 10828, A
45	225	2.3	1944	4	US-09-949-016-10929	Sequence 10929, A

ALIGNMENTS

RESULT 1

US-09-595-684B-31

; Sequence 31, Application US/09595684B

; Patent No. 6544766

; GENERAL INFORMATION:

; APPLICANT: Beraud, Christophe

; APPLICANT: Ohashi, Cara

; APPLICANT: Sakowicz, Roman

; APPLICANT: Vaisberg, Eugeni

; APPLICANT: Wood, Kenneth

; APPLICANT: Yu, Ming

; TITLE OF INVENTION: Human kinesins and methods of producing

; FILE OF INVENTION: and purifying human kinesins

; FILE REFERENCE: cytop036

; CURRENT APPLICATION NUMBER: US/09/595,684B

; CURRENT FILING DATE: 2002-06-24

; PRIOR APPLICATION NUMBER: 09/295,612

; PRIOR FILING DATE: 2000-04-20

; NUMBER OF SEQ ID NOS: 105

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 31

; LENGTH: 2662

; TYPE: PRT

; ORGANISM: Human

; US-09-595-684B-31

Query Match 3.1%; Score 301.5; DB 4; Length 2662;

Best Local Similarity 19.1%; Pred. No. 2.7e-11;

Matches 366; Conservative 330; Mismatches 676; Indels 549; Gaps 92;

Qy	1	MNKKSTLLATATAAIIIGSTVGTGVLGASIKYRGVNPQTGVSQGLIDSVAFKPSIA	60
Db	763	IQDKSEELHIITSEK---DKLFSVHVHKS---QGLLEIG-----KTKODLA	806
Qy	61	NFTSDYQSVKALLNGKTFDPKSEFTDFVSKPDLTNGRTVLLEIPKQYQVISEFSPE	120
Db	807	TTQSNYKSTDOEQFNKTLH-----MDFOKYQVILEE-----	839
Qy	121	DDKERFRGLPHLEKLEDGNIQAQSAKFIYLL-----PLDMPKAAAGQSYIVDKNFNL	175
Db	840	---NRM-----NQELVNLSEKQAQFDELGALKTSLSKTQELQKTEVQERLNM	889
Qy	176	--IIHPLSNFSAQSIKPL-ALTRSSDFIA-KLNFFNNOELMWVLEKFFDLEALKANIRL	231
Db	890	EQLKEQLENRDS---PLQTVREKTLITERKLOQ--TLEEVKTLTQEKDLDLQESLIQI	943
Qy	232	QTADFSFEKGNLVPFVYSFIRNPQNKQEWASDLNQDKTVRLYLRTFEFSQAATILKDY	291
Db	944	E-----RDQLKSDIHD-----TVNMNIDTQ--EQLRNALES	973

Dd	763	I QKSEELHIITSEK---DKLFSEVVHKESRV-----QGLLEIG-----KTODLA 806
Qy	61	NFTSDYQSVKALLNGKTFDPKSESEFTDFVSKPDEFLTNNGRVTVLBI PKKYQVWIFSEFSE 120
Dd	807	TTQSNYKSTQEQFNKTLH-----MDPEQKYKMWLE-----839
Qy	121	DDKERPRGLKBEKLELDGNIAQSAQTFIYL-----PLDMKAALQGYSYIVDNFNVL 175
Dd	840	--NERM-----NOEIVNLSKBAQFDPSSGALKTELSYKTQLOEKTREVOERLNM 889
Qy	176	--I IHPLNSFAQSIKPL-ALTRSSDFIA-KLNOFNMODELWYLEKFPDLKALKANIRL 231
Dd	890	EQLKEQLENRDS---PLQTVREKTLITIEKLOO--TLSEVKTLTQEKDDLKQLOESLQI 943
Qy	232	QTADFSPEKGNLVDPPFYISFIRNPONCKEWASDLNQODKTVRLYLRTFESPOAKTIL--- 288
Dd	944	ERDOL---KSDIHD-T-VNMNIDTQEOURLNALESUKQHQTINT-LKSKISEEVSRLHWE 998
Qy	289	-KQYKQDFTPLASI-----DLKANGTSLPAN-----ENDL---319
Dd	999	ENTGETKDFEQOKMVGIDKKQDLEAKNTQTLTADVKNDEIIEQORKIFSLIQEKNELOQM 1058
Qy	320	-----KOOLDVDLLD-----VSDYFGOSETIT---SNSQVQVPVPSERS 356
Dd	1059	LESVIAEKBQKLTDLKENIEMWTIENOHELRLLGDELKQOEIVAQEKHAIKKEGELSRT 1118
Qy	357	LKORVKFKQOQKPRIKFSFLYEDALSFYSQJQELVSKPNSIKOLVNATLARNLFRSIG 416
Dd	1119	C-DRLAEVBEKLEKSOOLQEQOOLLNVQEEMSEMQKINEIENLKNB-LKNKELTLE 1175
Qy	417	KYNFLDLDLASHLYYFLVSKAKIKOSITPKLFIELPTKISKSSILGDQBNKTIPLFE 476
Dd	1176	HMETERLELAQKLNENYEVKSITKERVKLEL-----OKSEF 1213
Qy	477	KEVTFKLDNPRDVEIEKAFGLLYPGVNEELQAKQASFEKSKKGLKEFPOQKEEN 536
Dd	1214	TERDHLRGYTHEIE---ATGL---OTKEELKIAH-----IHLKEHQETIDELRRSVSEK 1261
Qy	537	SKAINNOBGLDEEDNITERLPENSPIYOQOENAGLQASDPKPYMIDVQNQRYLAKSQI 596
Dd	1262	TAQIINTODLEKSHT---KLQEEIPVLHBEQEL-----LPNVKVSQTQETMNELELL 1311
Qy	597	QELIKAKDYTKLAKLNNRHTYNTLSLKQOLFQVNPRIPS-SDIEKAKFVLDTKEKNK 655
Dd	1312	TEQSTTKDSTTLARIEMER-----LRINEKFOBSQBEIKSLTKERDNLKTIKEALEV-K 1364
Qy	656	YWOIYSSASPQFNKWSLFGYRYVLLGLDPKQTIHE-LVKLGQKAGLOFEGYENLPSDEN 714
Dd	1365	HDOL-----KEHIRETLAKIQBSQSKQ-----BQSUN 1391
Qy	715	LEDLKNIRIKTPFSKDNKFLSLLD*NNYVDGEIKAPFEGLPFLPKELRNRRSSGGS 774
Dd	1392	MKEKN---ETTKIVSEMQFPK-----DSALLRIEIM-LGLSKRLQESH-----1435
Qy	775	QNSNSPWEQEIISQPKQNLNQNQOLAQFSTKIWE---KIIGDENEPDQNNRLOYKLLKD 831
Dd	1436	-----EMKSVAKEKDDLQRLQEVLSQESDLQAKENIKIYAKHLETEBELKVAHCCLKE 1488
Qy	832	LOESWINKTRNL-----YTYLGDKLKVKPKNNLEAK-----FROJSLQLELL 875
Dd	1489	-QEETINELRNLSEKETEISTIQKQEAINDLKNKIQEIYEKBEQOJNTKQISEQVENV 1547
Qy	876	TAF-----VTSAAALNNNNYQDSGAKSTIIFEIABLDPKVKKEKGVADV 920
Dd	1548	NELKQFKEHRKAKSALQSIKESKMLELNRLOESQEEIQIMIKEEM-KXVQEAQIER 1606
Qy	921	YOLKFHYAIGFDNAGKFNQEVIRSSRTI-YLKTSGSKSLEAD--TIDQNLQAVKNAPL 977
Dd	1607	DQK-----ENTKEIVAKKESQBEKEYQLKMTAVNTEQKWCSEIHLKEQFQTKL 1658
Qy	978	GLQSFYLDTERFGVQKLAATSLAVQH-----KQEKTLPKLNNDGYTLIHLKXK 1028
Dd	1659	NLEN--IETENLRLLQIILHENLEENMRSYTKERDRLRSVEETL--KVERD-QUKENIRE 1711

RESULT 3

RESULTS 3
IIS-09-949-016-6468

US-09-949-016-6468 : Sequence 6468. Application US/09949016

; Sequence 6468, App
: Patent No 6812339

; Patent No. 6812339
; GENERAL INFORMATION.
;

APPLICANT: VENTER, J. Craig et al.

APPLICANT: VENTER, J. CRAIG ET AL.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE
 ; TITLE OF INVENTION: METHODS OF DETECTION AND USES THEREOF

;
 TITLE OF INVENTION: WITH
 THE REFERENCE: C1001307

; FILE REFERENCE: CL001307
; CURRENT IDENTIFICATION NUMBER: IIS/00/949 016

; CURRENT APPLICATION NUMBER: US/09/944

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

```
; SEQ ID NO 6468
; LENGTH: 1979
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6468

Query Match      2.9%; Score 286.5; DB 4; Length 1979;
Best Local Similarity 19.0%; Pred. No. 1.9e-10;
Matches 373; Conservative 345; Mismatches 682; Indels 563; Gaps 99;

QY 22 FGVVGLASVYGVNGVPTQGVISQGLSDVAFKPSIANFTSD-----YQSVKKAL 73
Db 164 FGDIISSQGEIN-RLSNEVSRLESEVGHWRHTA-QTSKAGQTDNSQDSICKLQNIKEI 221
QY 74 LMGKTPDKPSSEPTDFVSPDFITN-NGRTVLEIPKKYQVWISSEFDEKFRFLGPHL 132
Db 222 KQN-----RSQETDDHQHMSVLQNAHQKQLTEISRRHREELSDY-----EERI----- 265
QY 133 KEKLEGNIAQSNATKIYLLPLDMPKAAALGOYSYIVDKNFNNLIHPLNFSQAQSIKPLA 192
Db 266 -BELE--NLQOQSGV--IETDLSKI-----YEMQKTIQVLIQIEKVE--STKKMEQLE 312
QY 193 LTRSSDFIAKLNQFN-----QDELWYLEKFPDLEALKANIRLQ-----TA 234
Db 313 -DKIDINKKLSAENDRIILRREQQLNV--EKROIIECE-NLKLECSKLQPSAVKQS 368
QY 235 DFSFEKGNLV-----DPF-VYSFIRNPQNKEMASDLNQD-----QKTVRILYLRTEFSP 282
Db 369 DTWTEKERILAQASAVEEVRFLQQAALSDAENEIMRLSSLNQDNSLAEDNLKMKRIEVL 428
QY 283 QAKTILKDYKDETFLSIDLK-----ASNGTSLPANENDLKQDLDVDDLLDVSDY 333
Db 429 KEKSLLS--QEBELQMSLLKLNNYEVIKATATRIDISLDSSELHDLRLNLE----- 477
QY 334 FGQSETITTSNQVQVPASERSLKDVRVKFKDQKQPRIEKFSLYEYDALSFVSQLOEVL 393
Db 478 -----AKEQLNQSIS-EKETLIAEIEELDRQNEBATKMILIKOQL 518
QY 394 SKPNSIKDLVNATLARNLRFSLGKYNFLFD---DLASHLDYFPLVSKAKIKOSSITKULF 450
Db 519 SKQNEGDSIISKQLDNDEKKRVHQLBDDKMDITKELD---VQKELIQSEVA--- 570
QY 451 IELPIKISLSKSSILGQDQENIKTLPEKEVTFKLDNPRDVEIEKAFGLLYPGVNEELEQAR 510
Db 571 -----LNDLH-----LTKQLEDKVENL-----VDQLN 593
QY 511 KAORASFEKSKKGLKEFSQOKEENSKAINNQ--EGLBEDDN-----ITER----- 555
Db 594 KSQESNVISQKENLELKEHIRQNEBELSRIRNELMQSLNQDNSNFKOTLLKEREAEVRN 653
QY 556 LPEN-SPIQVQBNAGLGASPDKPYMIKDVQNRYYLA---KSQIQELIKAKDYTKLAK 610
Db 654 LKQNLSELEQNENLKKVAFDVK-----MENEKLVLACEDVRHOLEECLAGNQLSLEK 707
QY 611 LLSNRHTYNISLRLKEQLFDVNPRISSRDIIEKAKFVLDKTEKNKYQWISYASSPVFQNK 670
Db 708 -----NTIVETLKMEGEIEAEELCWAKRLLEA-----NKYEKTIELSNA-RN- 751
QY 671 WSLFGYVRYLLGDDPKQTIHE-LVLKQKAGLQF-----EGYENLPSDF-----NLE 716
Db 752 -----LNTSALQLEBEHLIKLNQKQMDIEAELKQNIQOMDTHKETKDVLSSSLE 801
QY 717 DLKNIRIKTPLFSQDNFKLSLDDFNYYDGEIKAPEFGLPLPLPKELRRN----- 767
Db 802 EQQL---TQINKKEBIFIEKERSKQLEELDK-----YSQALRKNELIRQTIE 850
QY 768 SSSNGSQNSNSPWEQIEISQKDO-----NLSNQOLA-----QFST----- 805
Db 851 KORSLSGKMEENHQLQELERLEREEQSRAPVADPKTLDSTVTELASEVSQLNTIKEHLEE 910
QY 806 --KIWEKIIGDENEFQNNRQYKLLKDLQESWINKTRDNLWYTLGDKLVKPKQNL 863
Db 911 EIKGHQKIIEDNQ-----SKWQ--LLQSLQEQ--KKEMDE--FRYQHEQMNATHQLFLE 960
```

RESULT 4

US-09-949-016-7404

; Sequence 7404, Application US/09949016


```

Qy 1531 EYEKKSPKIAK--GWTNHPHQKEYPNPQKLPENYINLVLPNQPKVTLYNSSD----- 1582
Db 1793 EANAALDSASRLTEQLDQVYKQBIELKFKQNLQRMQLDDV--QKKLMGLANSSEKGVKDV 1850
Qy 1583 FITNLV-----EPEGSDR-----GSGTKLQVIOQVNNYADMGSAYLFWY----- 1626
Db 1951 LMRNLFIGHPHTPKNQREHVLRLMGSILGVRREMEQLFHDDQGGVTRWMTGMLGGGSKS 1910
Qy 1627 --DKNIITNQNVNTANTADVFIDKVDKELENTKLIAPNIT 1665
Db 1911 VPNTPLRPNQGVVNSSEFSELVKFL-ETESHPSIPPPKLS 1950

RESULT 5
US-09-949-016-6507
; Sequence 6507, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6507
; LENGTH: 3259
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6507

Query Match          2.8%; Score 276; DB 4; Length 3259;
Best Local Similarity 18.8%; Pred. No. 2,1e-09;
Matches 310; Conservative 240; Mismatches 505; Indels 590; Gaps 76;

Qy 106 IPKKYQVVVISFSPEDDKERFRLGLPHLKEKLEDG-----NIAQATKF 148
Db 1738 LSKKQFSLMSKDSLSEVQ-----DLKHQIEDNVSKQANLEATEKHNDQNTNVTBEQTQS 1792
Qy 149 IYLLPLDMPKAAALGOYSYIVDKNFNNLIHPLSNFSAOSIKPLALTRSSDFTAKLNQNN 208
Db 1793 I-----PGTEEQDSL-----SMTRPTCSSEVSFAKSANPAVSKD-----FSS 1831
Qy 209 QDELWVYLEKPFDFLEALKANIRLOTADPFSFKGNLVDPFVYSFIRPNQKQEWASDLNQD 268
Db 1832 HDEINNVYLOQ---IDQLKERI---AGLEEEK-----QCNKEFSQTLNE 1869
Qy 269 QKTVRLYLRTFSFOAKTILDKYKQDFTFLSSIDLKASNGTSL-----FANENDLKDQLD 324
Db 1870 KNTLSQISTK-DGELK-MLOSEVTKMNLNLNQIQEELSRVTKLKETAEEKQDLEERLM 1927
Qy 325 VDLLDVSDFYGGQSETITS-----NSQVKPV-----PAGERSLKORVKPKDQO 368
Db 1928 NQLAELNGSIGNYQDVTDAIQKNELLESEMKNLKKCVSELEEEKQQLVKETKYVESEIR 1987
Qy 369 KPRIETKF--SLYEVDAISFYSOLOELV-SKPNISIIDLNV-----ATLARNJR---- 412
Db 1988 KEYLEKIQGAQKEPGNKSHAKELQELLKQOEKVQLOKQICIRYOEKISALERTVYKALEF 2047
Qy 413 -----FSLGYNFLFDDLASHLDYFVLVSAKAKIKOSSI 445
Db 2048 VQTESQKDLKITKENLAQAVEHRKKAQAELASFVKVLLDDTQSEA-ARVLADNLKLKKELO 2106
Qy 446 TKKLFIELPIK---ISLKSSILGQGPENIKTLFEKEVTFKLDNFR--DVEIEKAFGLLYP 500
Db 2107 SNKESVKSQMKOKEDLERRLEQAEAEKHLK--KKNMOEKLDAALRRKVLHEETIGETQV 2164

```

Mon Nov 21 16:44:20 2005

Qy 1277 -RKSEELTYNPKDFNPLNLTAKHRLTSLNVLNNDPNYKIEDLVKIKNEAGDHL--- 1332
Db 3121 HRKSDPELREPOQSFEAQ-----QQLCNT--ROEVNELRLLBEERDORVAE 3168
Qy 1333 -AFSLRANNIKRL-----MNTPI 1349
Db 3169 NALSVABEQIRLREHSEWDSRTPI 3193

RESULT 6
PCT-US93-07261-16
; Sequence 16, Application PC/TUS9307261
; GENERAL INFORMATION:
; TITLE OF INVENTION: PREMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John H. C. Blasdale
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07261
; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasdale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0288K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1663 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; STRAIN: Malayan Camp
; PCT-US93-07261-16

Query Match 2.8%; Score 275.5; DB 5; Length 1663;
Best Local Similarity 20.5%; Pred. No. 8.3e-10;
Matches 332; Conservative 227; Mismatch 578; Indels 485; Gaps 82;

Qy 74 LNKG-----TFDP-----KSEFTDFVSKFDFTLNNGRTVLETPKYYQVVISFES 118
Db 389 VNGKKLSTFNPFKRRNKLKERKQELHFKKNY-----KKYQKLE--- 431
Qy 119 PEDDKERFRGLGHLEKLE--DGNIAQSAKTIYLLPLDMPKAALGOYSYIVDKNFNNLI 176
Db 432 -----REKRNPDGE-----PLNTPBIHVIRSDLMKXENKSA 465
Qy 177 IHPLSNFSASQIKPLALTRSSDFIAKLNQFNQ--DELWVYLEKFPDLEALKANIRLOTA 234
Db 466 GHP--FKYQPTKGLKEYES-HVSKDYQLEHEPPTKLPY-EKGHSREYQLDHEPPTK 520
Qy 235 DFSFEKGNLVDPF-VYSFTRN--PQNK-----EWASDLNQDQKTVRLYLRTEFSPQAKT 286
Db 521 LPEYEGHVSREYQLDNEVRDELPEYEGHVSREYQLD-NEGFTLKEYDQTELAKGDI 579
Qy 287 ILKDYKVKDETFLSSIDLKASNGTSLFANENDLKOOLD-----VDLLD 329

Db 580 TNKPHEVDEYDQTEL-----AKGKOITNKPHEVDEYDQSELAKGDKDITNKPHEVDEYD 635
Qy 330 VSDYFGQSEITITSNQVFPVPSRSLKDRVKFKDQOKPRIEKFSLEYEIALSFYSOL 389
Db 636 QTELAKGKEVTNKPHEHLE-----EYNETDLAKGKEVTNKPHEHLE-----ESVDEYDQ-SELAKG 686
Qy 390 QELVSKPNSIKDILVNATLARNLRFSLGKYNFLPDDIASHLYFLVSKAKIKOSSITKKL 449
Db 687 KDIITNKPHEVDEYDQ-----ELAKGKEVTNKPHEHLE-----EYNETDLAKGKE- VTNKA 737
Qy 450 FIELPIKISLKSILGQBPNIKTLPF-KEVTFK-----LDNFRDVEIEKAFGLLYPGVNE 504
Db 738 RENL-----EYNETDLAKGKEVTNKAHENLEEYNETDLAKGKEVTNK-AHE 783
Qy 505 ELFOAKQAQASPEK--EKSCKGLKEFSQ-----KEENSKAINNOBLEEDD-----NI 552
Db 784 NLEEYNETDLAKGKEVTNKAHENLEEYNETDLAKGKEVTNKAHENLEEYNETDLAKGKEV 843
Qy 553 TERLPENSPIOYQOENAGLGASDPKPYMIKDVOQ--RYVLAKSQIOELIKAKDYTKLAKL 611
Db 844 TNKARENLE-EYNETDLAKG-----KEVTNKAHENLEEYNETDLAKGKEVTNKAR- 892
Qy 612 LSNRHTYN-----ISURLKEQLFDVNPRIIPSSRDIEKAFVLDKTEKNKYQIYSS 662
Db 893 -ENLEEYNETDLAKGKEVTNKAHENLEEYNETDLAKGKEVTNKAHENLEEYNETDLAKGKEV 944
Qy 663 ASPVFQNKWSLFCYVRYLLGLDPKQITHELVLKQKAGLQFEGYENLPSPDNLEDLKNIR 722
Db 945 KQYMKNE-----LQNKSGDGLKENAELKN-KELR 973
Qy 723 IKTPFQSKDNFKLSLLDFNNYDGBIKAPEFGLPLPKELRRNSNSGGSQNSNSPWE 782
Db 974 NK-----GSDGLK-----ENAEIK-----NKEL-RNKSGDGLKEN----- 1002
Qy 783 QEIIQPKDQNLNQDLAQFSTKIWEKIIIGDENEFQDNRLQYKLLKDLQESWINKTRD 842
Db 1003 -----AELKNKELRNK-----GSEG-LKENAELKNKELR-----NKGE 1035
Qy 843 NLYWYTLGDKLVKPKNLEAKFRQISN-----LOELLTAPYTSAAALNNNNYQDSGAK 897
Db 1036 GL-----KENAELKNKELRNKGEGLKE-----NAELKN--KELQNGSE 1073
Qy 898 STIIFEBIAELDPKVKKEKVADVVYQLKFHYAIGFDDNAGKPNQEVIRSSRTIYLTSGK 957
Db 1074 G-----LKENAELKNKELQNGSE-----GLKENAEQKNKE-----LQNGKS 1110
Qy 958 SKLEADTIDQLNQAVKNAPIGLQSLQFYLDTERFQVQKLATSLAVQHQKQKTLPKKLND 1017
Db 1111 EGLKE-----NAELKNKELR-----NKGSDGLKENAELKNKELRNK-GSD 1149
Qy 1018 GYTLIHDKLKPVIPOISSPEKDWFGKLNQSONVNVSTFGSI-IESPYFSTNFOE 1076
Db 1150 G-----LKENAELKNKELRNKSGSLKENVYTNNDLK 1181
Qy 1077 DADLDQDQDSDROGNNSLDNQEAAGLLKQKLAIIILGNQFIQYQOONKBEIIEINVEKV 1136
Db 1182 NNDIQNDLSNKMKNKELLNKDINSKDMKNKELLNKDLSNEDMKNELLNKDIRNKDLK 1241
Qy 1137 SELSFVRVEFKLAKTLBNGKTIIRVLSDBETMBLIVNTTIEKTP-EMSAVPEVDFTKWEOY 1195
Db 1242 SIGNEQONTGLKNTPSKQOONTGLKNTPNEROQONTGLKNTPSKQOONTGLKNTPSKQO 1301
Qy 1196 D-----PRTPLAAKTFVLKFKDQIPVDGSGNISDKWLASIPLVIHQOMLRSLPVVKTIREL 1252
Db 1302 NTGLKNTPNEROQONTGLK-----NTPSEGOQNTG-----LKNTP-LEGQONT 1342
Qy 1253 GLKTEEQOQOQO-----QOQOQOQOQKAVRKEEBELETYNPKDEFNIIINPLTKAHLT 1304
Db 1343 GLKNTPSKQOONTGLKNAANKQOONTGLKNAANKQO-----NTGLKNTPSKQOONT 1394
Qy 1305 -LSNLVN-----NDPNYKIEDLVKIKNEAGDHLAFSLRANNIKRLNMTTITPADI 1354
Db 1395 GLKNTPNEROQONTGLKNTPNEROQONTG-LKNTPSKQOONTGLKNTPNEROQONTGLK----- 1449

QY 1355 NPFYNNEDWRSIDKYLNNKG--NVASHQOQAAG-----GNQSGSLQRLNKNITKPT 1405
Db 1450 -----NTASKGQONTGLKNAPNERQONTGLKNTPEGOQNTGL-----KN----- 1489
QY 1406 FTALIALKDRNNTNLSNYSYDK-----IIMIKPYLVERSIGVPMSTGLDGYIGSOTKD 1460
Db 1490 -TPS-----EGQONTGLKNAANKRQONTGLKNTPE-----SEGQP--NTGLKNTPEGOQNT 1537
QY 1461 G-----TSSSSQ-----RGFODDFIQAALGLKNTYHGLSLIRFFDPCNELAKIKDASNK- 1512
Db 1538 GLKNTPEGOQNTGLKNTPEGOQNTGLKNTPEGOQNTGLK-----NTPNEGOQNTDLKKA 1595
QY 1513 -KGEKLLKSYDLFPKNLYNEYEKSP-----KIAGWTNIHPDQ--KEYPNPNQKLP 1561
Db 1596 SKGOQNDLNDGLKHPNQGQKHTELNNKLNKPKPTDGLKNVYKDDLSNDESSDNEKSK 1655
QY 1562 EN 1563
Db 1656 KN 1657

RESULT 7
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064, 964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055, 779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 2.8%; Score 275.5; DB 3; Length 10182;
Best Local Similarity 18.5%; Pred. No. 1.2e-08;
Matches 375; Conservative 319; Mismatches 758; Indels 573; Gaps 93;
QY 74 LANGKT-----PDPKSSEBTFVSKFDFLTNNGRVLEIPKPYQVVISFSPEDDKERF 127
Db 3189 LNMGTQESINNYTKRREAQNTASSADTTINNGDASIE-----QITEN----- 3231
QY 128 LGPHLKEKLEDGNIAQSATKFIYLLPLDMPKAALGOYSYIVDKN-----FNNLIH 178
Db 3232 -----KIRVEATNALNEAKHLTAUTTSIKTEVRKLSRRGDTNNKKPSSVAYNN-TIH 3285
QY 179 PLSNFAQSQIKPIALTRSSDFTAKLNQFNQDELWVYLEKFFDLEALKANIRLOTFADFSF 238
Db 3286 SLOSEITQT-ENRANTIINKPIRSVEEVNHALH-----EVNQLNRLT----- 3327
QY 239 EKNGLVDPFVYSFIRPNQKQWASDL-NODQKTVLRLTPEFSPOAKTILKDYKYKDET 297
Db 3328 DTINLLQP-----LANKESLUKEARNLESKINETVQTDGTOQS-----VENYK----- 3371
QY 298 FLSSIDLKASNGTSL---FANENDLKDQ-----LDVDDLVDVDFGQSETTITSNQ 346
Db 3372 ---QAKIKAQNESSIAQTLLINNGDASDQEVSTIEKLNKLSLTN-----SINHLTVNKE 3424
QY 347 VKPVPASERSLDRVRKFKDQKPRTEKFSLEYDALSYSQOLQELVSPKNSIKDLVNAT 406
Db 3425 --PLETAKNQLQANI-----DQKPSD--GMTQOSVQSVYERKIQEAKDKINSI-----NNV 3471
QY 407 LARNLRFSLGKYNFLFDDLASHLDYFVLYSKAKIKOSSITKKLFIELPIKISKSILGD 466

Db 3472 LANNPDVNAIRTN-----KVFTEQINNEL-----TOAQQLTVD 3505
QY 467 QEP--NITKLPK-----EVTFKLDNFRDVEIEKAFGLLYPGWSELEQAR-K 511
Db 3506 KOPLINAKTALQOOSLDNQPSITGTMTEATONTYNAKRQAEQ-----VIQANKIENAOFS 3561
QY 512 AQRASFEKEKSKGLKEFSQ-----QKEBNSKAINNOEGLEEDDNTITERLPENSPI-- 562
Db 3562 VQOVSDEKSKVEQALSELNNAKSAALRADKQELQQAQYQ-----LIQPTDLNNKKPASITAYN 3618
QY 563 -QYQENAGLGAS-----PDKPYMIKDVONQRYLYAKSQIOELIKAKDYTKLAKLS 613
Db 3619 ORYQOFSNELSTKNTDRILKEQNPVSADVNN-----ALNKVREV--QOKLINEARALQ 3671
QY 614 NRHTYNISLRLEQDFVNPRIPISS-----RDIEKAKFVLQKTEKN 654
Db 3672 NKEDNSALVRAKEQLQQAQVDPSTEGMTQTKDDYNSKQAAQOEISKAQOVIDNGDAT 3731
QY 655 KYWQIYSSASPVFQNKWSLFGYYRYLLGLDPKQTHIELVKGQKAGLOFEGYENILPSPDN 714
Db 3732 T--QOISNAKTNVERALEALNNAKTGLRADKEELQNAVYNLTQ-----NIDTSGKTPASIR 3785
QY 715 LEDLKNIRIKTPLFSQKDNFKLSLLDFNYYD-----GEIKA--PEFGLPLFLPKELRR 766
Db 3786 KYNEAKSRIQTQIDSAKNEANSILTDNPNQVSVQVTAALNKIKAVQPELDKATAMLNKEN 3845
QY 767 NSSNSGGSG-----NSNSPME---QEIIISQFQDNLSNQDQLAQFSTKIMEKII--GDE 815
Db 3846 NNALVQAKQOQLOQIVNEVDPTQGMTTDTANNYKSKKREADEIQKA-----QOILNNGDA 3900
QY 816 NE---FDQNNRLQYKLLKDLQESWINKTRDNLYWYLGDKLVKPK-----NNLEAKFRQ 867
Db 3901 TEQQTINETNRVNOAI-----NAINKAKNDL---RADKSQLENAYNOLIQNVDNNGKK 3950
QY 868 ISNLOELLTAFYTSAAALSNNNNYODSGAKSTIIIEETAEILDPKVKEKGA---DVOYL 923
Db 3951 PASIOOYQAA---ROAIEQYN-----NAKSE-AHQILENSNPVSNEVAQALQKVEAVOL 4001
QY 924 KPHYAIGFDDNAGKFNQEVIRSSRTIYLTSGSKLEADTTIDQLN-----QAVK 973
Db 4002 KVNDAILHLQN---KENNSALVTAKNLOQSVNDQPLTTGTMQDSINNYEAKRENAQSAIR 4059
QY 974 NAPLGLQ-----SFYLDTERFQVFOKLATSLAVQHKQKEKTLUPKKLNNDGYTLIHDKLLK 1028
Db 4060 NAEAVINNGDATAKQISDEKSKVEQALA-----HLNDAKQQLTADTTTEL 4103
QY 1029 PVIPOISSPEKDFEGKLNQNGQSONV---NVSTFGSIIIESPYFSTNPFQEDADLDQDQ 1085
Db 4104 QTAVQ-----QLNRGDTNNKKPSINAYNKAIQSL----- 4134
QY 1086 DDSRQGNNSLNDQEAAGLLKQKLAILLGNQFIQYQOQNDKEIFEIINVEKVS-----L 1139
Db 4135 --ETQITSAKDNANAVIQPIRTVQEVNNALQOVNQLAQOLTEAINQLPLSNNDALKAA 4192
QY 1140 SRVFEFLAKLTLEDNGKTIKIRVLSDETMSLIVNTTIETKTPMSAVPEVPTKVEQYDPT 1199
Db 4193 RLNLKINKINQTVQTDGMT-----QOSIEAYQNAKVAQNESNT 4230
QY 1200 PLAAKTKFVLKFKDQIPVDGSGNISDKWLASIPLVIHQMLRLSPVVK--TIRELGLKTE 1257
Db 4231 ALA-----LINNGDADEQOITTEDRVNQOITNLTOAINGLTVYKPELET- 4275
QY 1258 OQOQOQOQOQOQOQKAVRKEEBELETYNPKDEFNINPLTKAHLTLTSLNVLNNDPNYKI 1317
Db 4276 -AKTALQNNIDQVPSTDGM-TQOSVANYNQLQ-----IAKNEINTINNVLANNP----- 4323
QY 1318 EDLKVIK-NEAGDHQALFSL-RANN-----IKRLMNTPI---TFAD---YNPFF 1358
Db 4324 -DVNAIKTNKAEARISNDLTQAKNQLQVDTQPLEKIKRQLODEIDQGTNTDGMTQDSVD 4382
QY 1359 YNEDWRS-----IDKYLNNKGVSSHQOQAAGNQGSLIQ----- 1395

Db	4383	NYNDSLSAAIEKGVKNLKRNPVTEQVKESVANAQQ----	VIQDLQARTSLVPDKTQL	4439
Qy	1396	-----RLNKNIKPFTFPALJALKDRNNTL	SNYSOKIIMKPKYLVERSIGVPMWSTGLD	1450
Db	4440	QEAKNRLSNSINQOTDTG-----	TQDSLNNYNDK--LAKARONLEKI-----	SKVLG 4486
Qy	1451	GYISEQTKGTSSSSQKGFQDFICALGLK-----	NTEYH-----	GK 1489
Db	4487	GOPTVABIRQNTDEANAHK-----	QALDTARSQTLNREPYNHNNESHNLNAQKON	4539
Qy	1490	LGLSIRIFDPGNELEKIK-----	DASNKKGEEKL-----	KSYD- 1523
Db	4540	FKAQVNSAPNHTLETIKKADTLNQSWTALS	ESIAIDYENQKQOQENYLDASNNKRDYDN	4599
Qy	1524	---LFRKNYLN-----	YEKKSPIAKGWTNIHPDQKEYPNPKLPENYLNLV--	1568
Db	4600	AVNAAGTILNQTPGTSADVIDQCAEDVKRTK	TALDGNORLEVAQOQAL--NHLNTLND	4657
Qy	1569	LNQPKWTLNNSDFITLNFVEPEGSDRGSGTKL	QVIOKOVNNYAD--WGSAYLTFWY	1626
Db	4658	LNDAQROTLDTINHSPNINSVNOAKEKANTVNT	MTMTQTKOTIANYDDELHDGNYINADK	4717
Qy	1627	DK-----NIITNOPNITANIADVIFKDVKE	LENDTKLIAPNITQWPNISGSKEKFPK	1681
Db	4718	DKDAYNANNAQLINQSDANQAQDPAEINKVT	QRVNTTKNDLNGNDKLAERKRDAN	4777
Qy	1682	TVFPG-----NWNEN-----	SSMNSQAQPTTWEKIREGPAQALKSSFD-----	Q 1722
Db	4778	TTIDGLTYLNEAQRNKAQENVGKASTKTNITS	QLQDYNQL--NIAMQALRNSVNDVNVK	4835
Qy	1723	KTRFVLTTNAPLWPKYGP-----	LGFGNGP-----	NFKTQDMRLVFQNDNDQIAA 1769
Db	4836	ANSYNINEDNGPKAEAYNOAVTHAOTLINAQ	SNPEMSRDVNVNQTAQVNTAHQNLHGQOKL	4895
Qy	1770	LRVBOQ-----	DRPEKSSDKOKWIKFKVVIPEEMFNSGNIR	1808
Db	4896	EOAQSSANTEIGNLNLNTQKAKEK-----	ELVNSKQTR 4930	
RESULT 8				
US-09-949-016-11433				
; Sequence 11433, Application US/09949016				
; Patent No. 6812339				
; GENERAL INFORMATION:				
; APPLICANT: VENTER, J. Craig et al.				
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED				
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF				
; FILE REFERENCE: CL001307				
; CURRENT APPLICATION NUMBER: US/09/949,016				
; CURRENT FILING DATE: 2000-04-14				
; PRIOR APPLICATION NUMBER: 60/241,755				
; PRIOR FILING DATE: 2000-10-20				
; PRIOR APPLICATION NUMBER: 60/237,768				
; PRIOR FILING DATE: 2000-10-03				
; PRIOR APPLICATION NUMBER: 60/231,498				
; PRIOR FILING DATE: 2000-09-08				
; NUMBER OF SEQ ID NOS: 207012				
; SOFTWARE: FASTSEQ for Windows Version 4.0				
; SEQ ID NO 11433				
; LENGTH: 2733				
; TYPE: PRT				
; ORGANISM: Human				
US-09-949-016-11433				
Query Match				
Best Local Similarity 2.8%; Score 271; DB 4; Length 2733;				
Matches 308; Conservative 246; Mismatches 507; Indels 578; Gaps 76;				
Qy	106	IPKQYQVJISEPDDKFRFLGFLHKEJEDGNTAQATKFIYLLPLDMPKALGQVS	165	
Db	1212	LSSKQSLMSKDSLSEEVQ-----DLKHQIE-GNVSKQAN-----	LEATEKHNDQTN	1258
Qy	166	Y-----	IVDKVFNLIHPLSNFSAQSIKPLATLRSSDFAKLNQFNQDELWV	214

Db 2214 MSSLQNSRDRHANEDELKPKYDASLKEAQLKEQGL-----LNRE-----RDALLSETAFS 2265
Qy 1034 ISSSPKDWPE-GKLNQNGSQNVNVSTFGSIIESPY-----FS-----TNFQEDAD----- 1079
Db 2266 MNSTEENSLHLEKLNQQLSKDBQLLHLSQLEDVYNQVQSFKAWASIQNERDHLWNE 2325
Qy 1080 -----LDQCGQDDB-----ROGNSLONQEAAGLLKQKLAILLGNQFTQYYQ 1120
Db 2326 LEKFRKSEEGKORSAAQPSAPAEVOSLQKAMSLQNDRLDRLKE-----LKNLQQQYLQ 2380
Qy 1121 QN-----DKEIEFEIINVE-----KVSELSF 1141
Db 2381 INBEITELPLKALQEQYQDKTAKFQIMQBELKQENLSWOHELHQLRMEKSSWEIHERRM 2440
Qy 1142 RVBFKLAKTLIEDNG-----KTIRVLSDETMSLIVNTTIEKTPEMSAVPE-----VFD 1188
Db 2441 KEQYLMAISDKDQQLSHLQNLRELSSSSOTPLKVQYQORQASPETSPDGSQNLVYE 2500
Qy 1189 TKWVEQVDPPTPLAAKTKFVLKPK-----DQIPVDGSGNISDKWLASIPLVTHQMLRLSPV 1245
Db 2501 TELL-----RTQNDLSLKEITHQELRTIQQLNSNFSQLLEBKNTLSIQLCDTSQSLR----- 2551
Qy 1246 VKTIRELGLKTEQOQ-----QOQOQOQOQOQ-----QKKAV-RKEES 1281
Db 2552 -----ENQOHVGDLLNHCAVLEKQVQLQAGPLNIDVAPGAPOEKNGVHRKSDP 2600
Qy 1282 LETYNPKDEFNILNPLTKAHLRTLSNLVNDPNYKIEDLKVIKNEAGDHQ-----AFSLR 1337
Db 2601 EELREPOQSFEAQ-----QQLCNT--RQEVNELRKLLEERDQORVAENALSA 2648
Qy 1338 ANNIKRL-----MNTPI 1349
Db 2649 EQQIRRLHSEWDSSRTPI 2667
RESULT 9
US-09-710-279-2964
; Sequence 2964, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: F034800S
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 2964
; LENGTH: 5024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (5024)
; OTHER INFORMATION: variable amino acid
US-09-710-279-2964
Query Match 2.7%; Score 262.5; DB 4; Length 5024;
Best Local Similarity 19.6%; Pred. No. 3.4e-08;
Matches 392; Conservative 272; Mismatches 731; Indels 605; Gaps 97;
Qy 64 SDYQSV--KKALLNGKTFDPKSEFTDFVSKFDFLNTNGRTVLE-----IPKQVQVISEFS 118
Db 135 SDNESVNESYNIAE--PEKQHAFTALNNAKEIVNEQOATLDANSINQKAQAILTTKN 192
Qy 119 PEDDKERFLRG-----PHLKEKLEDG-----NIAQSAKTFIYLLPLDMPKALG 162
Db 193 ALDGEQLRAKENAQOEINTNLQTLDAQNSEKGLVNSQTRTEVASQI-----AKAK 246

Qy 163 QYSYIVDKNFNNLIHPLSNFSAQSIKPLALTRSSDFIAKLNQFNQDELWVYLEKFFDL 222
Db 247 ELNKVMEQ-LNNLI-----NGKNQMI-----NSSKFI--NEDANQOQ--AYSNAIASA 289
Qy 223 EALK--ANIRLOTADFPSEKGNLVDPPFVYSFIRPQNO-----KEMA-----SDLN 266
Db 290 EYLNKNSQNPELDVTYIEQAINN-----IN3AINNLNGEAKLTAKEDAVASINNLSGLT 344
Qy 267 OQKTV--RLYLRTRFSPQAKTILKDYKYKDTFSLSSIDL---KASNTGSLPANENDL- 319
Db 345 NEQKTKENQAVNGSQTRQDVANVLKSDKALDOSQTLRDLVNQNVHSTSYNFNEDSTQ 404
Qy 320 KDQLDVLDDVDYFGGQ-----SETITS-NSQVKPVPASERSLKDR----- 360
Db 405 KNTYDAINGSTYITGQHNSLNKSTIDQTSIQINTAKNDLHGAELQDRKGTANQEIG 464
Qy 361 -VKPKDQOKPRIKPSLEYEDALSFSYQQLQELVKPNSIKOLVNTATLAR-NLRFSLGKY 418
Db 465 QLGYLNDPQKSAEESLVNGSNTRSEVEHLNEAKSLNNAMKQLRDKVAEKTNVKQS---S 521
Qy 419 NFLFDDLASHLDYYFLVSKAK-----TKOSSITKLFIELPIKISLASS-ILGQ 467
Db 522 DYINDSTEHRGYDQALQEAENIINEIGNPTLNKSEIEQQLQQLTDAQNALQOGLHLEA 581
Qy 468 EPNIKLFEKEVTPKLDNFRDVEIEKAFGLL-----YFCVNEEL----- 506
Db 582 KNAIT---EIN-KLTALNDAQRKATENVQAOQTIPAVNQQLTLDRREINTAMQALRDK 636
Qy 507 --EQARKAQRASF--EKEKSKG-----LKEFSQOKEENSKAIN----- 541
Db 637 VQOQNVHQQSNYFNEDEQPKHYNDSVQAGOTIIDKLQDPTMKNKEIEQAINTTQT 696
Qy 542 ---NOEGLEEDDNITERLPEN-SPIYOQENAGLGASDPKPYMIKDVONQRYVLAKSQIQ 597
Db 697 ALSGENKLUHDOESTNRQIEGLSSLNTAQINAE-----KDLVNO----- 735
Qy 598 ELIKADYTKLAKLSNRHTYNISLRLKEQLPDVNPRISSRDIKAKFVL--DKTEKNK 655
Db 736 ---AKTRTDVAQLATAKEINSAMS-----NLRDGIQNKEDIKRSSAYINADPTKVTA 785
Qy 656 YKQIYSSASPVEQNKWSLFGYRYLLGLDP-----KQTHIELVKLGOKAGLOFEGVENLP 710
Db 786 YDQALQNAENI-----INATPNVELNKATIEQALSrvQQAQDLQDGVQQA 831
Qy 711 SDFNLEDLKNIRIKTPLFSQDNFKLSLLDPNNYDGEIKAPEFGLPLPLPKELRRNSN 770
Db 832 -----NAKQOATQTVNGLNSLNDGQKR--ELNLLI-----NSAN 863
Qy 771 S-----GGSQNSNPWE-----QEIIFQDKONLSNQDLAQFSTKIWEKIIGDEN 816
Db 864 TRTKVQEELNKATESNHAMEALRNSVQNVDDYVKQSNVYVNEQPEQHN---YDNVNEAQ 920
Qy 817 EFDQNR--LOYKLLKLDQESMINKTRDNLWYLYGDKLVKVPKNLEAKFQISNLQF- 873
Db 921 ATINNNAQFVLDKLAIERLTQTVNTTYDQALH---GTOKLIQDQQAAGTIGRLTSLNRP 976
Qy 874 -----LLTAFYTSAAALSNNWNYQDSGAKSTIIPETAEILD-----PKVKEK--V 916
Db 977 QKNAEVAKVTAATTRDEVN-----IRQEAATTLDTAMLGLRKSIKDKNDTK 1022
Qy 917 GADVYQLKPHYAIGFDDNAGKFNQBVIRSSSTIYLTSGKSKLBADTIDQLNQAVKNAP 976
Db 1023 NSSKYINEDHQQQAYDNAVNAQAVIDETQAT-----LSSDTINQLANAVTQAK 1072
Qy 977 LGLQSFYLDTERPGVFGQKLATSLAVQHKQKEXTLPKLLNNDGYTLIHDK-LKKPVPIS 1035
Db 1073 SNLHG---DTK-----LQHKDKSAKQTTAQIQ 1096
Qy 1036 SSPKDWPEKGLNQSQNVNVSTFGSITSEFSTNFQEDADLDQDQDSDROGNNSL 1095
Db 1097 -----NLNSAQKHMB-----DSLIDNSTRSTQVQHDLTBAQ-----AL 1129

Db 945 KDTWKONE-----LQKSGDGLKENAELKN-KELR 973
Qy 723 IKTPFSQKDNFKLSLLDFNNYDGEIKAPFGLPLPKELRRNSNGSGSNSPWE 782
Db 974 NK-----GSDGLK-----ENAEUK-----NKEL-RNKSGDLKEN-----1002
Qy 783 QEISQKQNLNQDLAQFSTKIWEKIIGDNEFDQNNRLQYKLLKDLQESWINKTRD 842
Db 1003 -----AELKNKELRNK-----GSEG-LKENAELKNKELR-----NKGSE 1035
Qy 843 NLYWYLGDKLVKPKNNLEAKFRQISN-----LQELLTAFYTSAALESNNWNYQDSGAK 897
Db 1036 GL-----KENAEUKELRNKSGSEGLKE-----NAELKN-----KELQNKSGSE 1073
Qy 898 STIIFEBIABLDPRKVEKVGADYVQLKFHYAIFGDDNAGKFNQEVIRSSRTTYLKTSGK 957
Db 1074 G---LKENAEUKELQNKSGSE-----GLKENAEQKNKE-----LQNKGS 1110
Qy 958 SKLEADTIDOLNOAVKNAPLGLQSLFYLDTERFVFOKLATSLAVQHOKKEKTLPKKLND 1017
Db 1111 EGLKE-----NAELKNKELR-----NKGSDGLKENAELKNKELRNK-GSD 1149
Qy 1018 GYTLLHDKLKPVPQISSPEKDWFEGLNQNGQSNVNVSTFGSI-IESPVFTNFOE 1076
Db 1150 G-----LKENAEUKELRNKSGSEGLKENVYITNDLK 1181
Qy 1077 DADLDQDQDSDRQGNNSLDNBAGLLKQKLAILLGNQFTQYQQQNDKEIEFBIINVEKV 1136
Db 1182 NNDIQKDLNKMKNKELLANKDISNKMKNKELLNKLNSNEDMKNKELLNKNKDLK 1241
Qy 1137 SELSPRVEFLAKTELDNGKTIIRVLSDETMSLLIWNITIEKTP-EMSAVEPDTKWEQY 1195
Db 1242 SIGMEQONTGLKNTPSKGQONTGLKNTPNERQONTGLKNTPSEGQONTGLKNTPSEGQ 1301
Qy 1196 D---PRPLAAKTKFVLKPKDQIPVDGSGNISDKWLASIPLVHQMLRLSPVVKTIREL 1252
Db 1302 NTGLKNTPNERQONTGLK---NTPSEGOQNTG---LKNTP-IEGQONT 1342
Qy 1253 GLKTEGOQOQOQ-----QOQOQOQPKKAVRKEBELETYNPKDFBNILNPLTKAHLRT 1304
Db 1343 GLKNTPSEGQONTGLKNAANKGQONTGLKNAANKGQO-----NTGLKNTPSKGQONT 1394
Qy 1305 -LSNLVN-----NDPNYKIEDLVKIKNEAGDHLQFLAPLRANNIKRLMNTPTTFADY 1354
Db 1395 GLKNTPNERQONTGLKNTPNERQONTG-LKNTPSEGQONNDLKNTPNERQONTGLK---1449
Qy 1355 NPFYFYNEDWRSIDKVLNKG--NVSSHQOQOAG-----GNQSGLIQRLNKNKIPET 1405
Db 1450 -----NTASKGQONTGLKNAPNERQONTGLKNTPSEGQONTGL-----KN-----1489
Qy 1406 FTPALIALKDRNNTNLSNYSKDIIMIKPKYLVERSIGVPMSTGLDGYIGSEQTKDGTSSS 1465
Db 1490 -TPS-----EQQONTGLKNTPN-----ERQ-----QNTGLKNTPSKGQONTGLKNA 1529
Qy 1466 SQKGFDQDFIALGLKNTEYHGLGLSIRIPDPGNELAKIKDASNNKKEELKLSYDLF 1525
Db 1530 PNERQ-----QNTGLKNTPSEGQONTGLK-----NSASKGQO-----1561
Qy 1526 KNYLNEVEKSPKIAKGWTHIHDPQKEYPNPQK 1559
Db 1562 ----NTGLKNTP--SEGOQN--NDLKNAPNERQO 1587

RESULT 11
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match 2.6%; Score 256.5; DB 3; Length 3696;
Best Local Similarity 18.2%; Pred. No. 5.5e-08;
Matches 415; Conservative 370; Mismatches 794; Indels 703; Gaps 102;

Qy 30 SKVKYRGVNP-----TQGVISQLGLSDVAFKPSIANFTSDYQ-----SVKKA 72
Db 995 SEAKNGINTINQDQPNVAKKNTKTIIEQKNEKKSAIAQTPDATTEBKQAVSAVQA 1054
Qy 73 LLNGKTFDPKSSEFTDFVSKF-----DFLTNNGRTVLEIPKTYQVVISSEFSPS-----120
Db 1055 VTNGITHINQANSDDVDQELSNAGIITHTVNVVQKQKQARQALIAKTNEKQASINDN 1114
Qy 121 ----DDKERFRUGFHLKEKLEDGNTAQSATKIYLLPLDMPKAAALGOYSYIVDKPNPNU 176
Db 1115 EGTIBEKQAIQSLANDAKNLADEQITQAA-----NQNVVDNAL 1152
Qy 177 IHPLSNFS-----AQSIKPLALTRSSDFTAKLNO---FNNQDELWVYLEKFFDLEALK 226
Db 1153 NTGISNISKIQNTFTKQARQOVNQKQOEKAEALNSTPHATQDEKODALTRLTQKETA 1212
Qy 227 ANIRLQTAADFSEKGNL-VDPFVYGFIRPNQKQEWASDLNQDQKTVRILYLRTEFSPQAK 285
Db 1213 LN-----DINQAOQNQVDTALTSGIQNIQ-----TQVNVVRKK-----QEAK 1250
Qy 286 TILKDYKYKDETFSSIDLKASNGTSLFANENDLKQDLQDVLDD-----VSDYP 334
Db 1251 TTINDIVQHQKQSIQNNDDATTBEEKEVANNLVNASQNVVISKIDNATTNNQIDGIVSD--1308
Qy 335 GGQS-ETITNSQVK-----PVPASERSLK-DRVKPKKDOQ---KPIREKFSLEYD 381
Db 1309 GQOSINAITPTDTSIKRNNAKNDIDIKAAADKKIKIQRINDATDEEIOEANKRIEAKIEAD 1368
Qy 382 AL---SFYSQLOELVSKPNSIKDLVNLATLARNLRFSL-----GKYNFLPDDLASHLDYF 433
Db 1369 NIQRNSTRDQVNE--AKTNGINKIENITPATTVKSEARQAVQNKANEQINHQTPTD---1423
Qy 434 LVSKAKIKQSSITK--KLFIELPIKISLKSqTLGDOEPNITKLPEKVT-FKLDNFRDVE 490
Db 1424 --ATNEEKQEAIRNVSAELARVQAQINAEHTTQG-----VKTIKODAITSLSRINAQVVE 1476
Qy 491 IEKAFGLLYPGVNEELEQAKQARASFEKEKSKGLKEFSQKQKEENSKAINNOEGLEED 550
Db 1477 KESARNATEQATQQTQTFINNNDNATDEKQVANNLVATKQ-----KSLDNINSLSNN 1531
Qy 551 NITERLPENSPIYOQENAG-LGASPDPKYMIKQVONQRYYLAKSQIOELIKAKDYTKLA 609
Db 1532 DV-----ENAKVAGINEIANVLPATAVKSKAKKDI-DQKLAQIQINQIQHTQATTEEKEA 1585
Qy 610 KL-----LSNRHTYNISLRLEKEQ-LFDVNPRIIPSSRDIEKAKFVLDTKEKNK 655
Db 1586 AIQLANQKSNEARTAIQNEHSNNGVAQKSNQSIHEIELVMPDAHKKSDAKSID-----NK 1641
Qy 656 YWQIYS--SASPVFQNKWSLFGYRYLLGLDLPKQTIHELVLKLGQKAGLQFEGEYNLPSPDF 713
Db 1642 YNEQSNITNTPTDADT-----BEKQALDKLKIADKAG--YNKVDQAQTNQ 1685
Qy 714 NLEDLKNIRIKTFLPSQKDNFKLSLLDFNNYVDGIBIKAPFGLPLFLPKELRRNSNSGG 773
Db 1686 QVSDAKTEAIDT-----ITNIQANVAKPSPARVELDSKPFEDLKRQINA--1728
Qy 774 SQNSNSPWEQEIISQFKD-----QNLSNQDQLAQSFSTKIWEKIIIGDNEFDQNNRL---824

Db 1729 TPNATEEKQAIQRLNGKREVKMLINQDR-----RDNVEQHKNIGLOE 1774
Qy 825 -----QYKLLKOLQESWINKT---RDNLYTYLGDKLKVKPKNNLE-AKFRQISN 870
Db 1775 LETIHANPRKSDALQELQTKFISQTELINNNKDAT---NEEKDEAKLLBISKNKTITN 1831
Qy 871 LOELLTAFYTSALSNWNYQDSAKSSTIIFBEIAELDPKVEKVGADVYQLKPHYAIG 930
Db 1832 INQAQTNQVNDNAKNGHNEIATIIIPATTIKDTAKTAIDKKAEOQVTI-----1879
Qy 931 PDDNAGKFNQEVIRSSRTIVLKTSGSKL-EADTIDOLNOAKVKNAPLGLOSFYLDTERF 989
Db 1880 INGNNDATDEEKAEARKLVEKAKIPEAKNITSUTEREVNGAKTN---GLEKIN-----1930
Qy 990 GVFOKLATSLAVQHKQKKTLPKLNNDGYTLIHDKLLKVPVPIQISSPEKDFEGK---1046
Db 1931 -----NQPSQTQKTNAKQ-----EINDKAQEQLI-QINNTPDATEEKEQAT 1972
Qy 1047 -----LNQSGSQNVNVTFGSI-----IESPYFSTNFQEDADLDQD 1083
Db 1973 NRVNAGLAQAIQINNAHSTQVNESKTSIATIKSVQPNVKKPTAINSLTQEAANNQKT 2032
Qy 1084 --GODDSRQGNNSLQDQAG--LLKQKLAILLGNQFIOYQONDKEIEFEIINVEKVSSEL 1139
Db 2033 LIGND-----GNATDDEKAAKQLVTKL-----NEQIKIHESQDQNDQVNVKAQAITAI 2083
Qy 1140 SP-----RVEFKLAKTLENGKT-IRVLSDETMSLIVNTTIEKTPMSAVPEVFT- 1189
Db 2084 KLINANAHRQDAIINLTNLAESKSDIRANQDA-----TTEKN---TAIQSIDDTL 2133
Qy 1190 -----KWVEQYDPRTPLAATKFKVLKPKQOI PVDGSGNISDKWLA 1229
Db 2134 AQARNNINGANTNALVDENLEDGKQKQRLVLSQTKQAK-----A 2175
Qy 1230 SPLVTHQOMLRSLPVVKTIRELGLTKTQOOOQOQOQOQKKA VKKEBELETYNPKD 1289
Db 2176 DIAQAIGQOR-----STIDQONQATTEEKOEAERLNQETNGVNDRIQAALANQWTD 2228
Qy 1290 EFNIL-----NPLTKAHLITLSNLVNDPNYKIED-----LKVKIN 1325
Db 2229 EKNNILETIRNVEPIVIVPKANEIRKKAABQTTLINQONQATLEEQIALGKLEEVN 2288
Qy 1326 BAGDQLAFSLRANNIKRLMNT-----PITFADYVPPFYNNEDWRS-IDKYLNNKG 1375
Db 2289 EA-LNQVSOAHNSNDVKTAEENNGIAKISEVHPETIIKRNAKQIEBQDAQSQIDTI--NAN 2345
Qy 1376 NVSSHQOQAGGQSGLIQRLN-----KNIKPETTPALIALKDRNNTLSNVSKI 1428
Db 2346 NKSTNEEKSA-----IDRVNVAKIDAINNITNATTQLVNDKNGSNTSIS-----2392
Qy 1429 IMIKPKYLVERSIGVPSWSTGLDYGISE-----QTKDGTSSSQQKGFQDQFIQ-- 1477
Db 2393 -QILPSTAVK-----TNALAALASEAKNNAIIDQTPNATAEKESEANNKVDRLQEE 2443
Qy 1478 -----ALGLKNTEYHGKLGLSIRIF---DPGNELAKIKDASNKKGE-----1515
Db 2444 ADANILKAHTTDEVNINKQAVQINAVQVEIKQNVKNQNLQFDNQKIIIENTPDAT 2503
Qy 1516 -EKLKSYDLFKNYLN-----EYEKKSPKIAKWTNHPDQ 1550
Db 2504 LEEKAEANRLLQNLVLTSTDEIANVNDHNNEVDQALDKAPKTEIIVPQVSKRDLVNAIQ 2563
Qy 1551 KVPNPNQKLPENY-----LNLVLNQPKVTL-----YNSDPI 1584
Db 2564 EAPNSQTIHQENQEBATNEEKEALNKINQLNQA-KVNIDQAQSKVDVSAKTSIQDI 2622
Qy 1585 TNLVFEPEGSDRG-----SGTKLKQVIOK-----1608
Db 2623 EQIQPHQTKATGRHRLNEKANQOQSTIATHPNSTIEERQEAQSLQEVKKAIADKDG 2682
Qy 1609 QVNNYADGSAVLTFTWYDKNI-----ITNOPNVITANI-----ADVFI-KOVKELEDWT- 1657

Db 2683 QTND-----VEKTVVNGIAEINILPATTVKDKAKADVNAEKEQKNLQINSN 2730
Qy 1658 -----KLI-----APNIQWPNISGSKEK-----FYK 1680
Db 2731 DEATTEBKJVASDNLHNHVETTNQATBEDAPDTNQ-----VNVKKNKGIGTIRDIQPLVVKK 2786
Qy 1681 PTV---PFGWNESENSMNSQAOTPTWEKIREGFAQALKSSPQO-----KTRTFVLTT 1731
Db 2787 PTAKSKIESAVEKKKTEIN-QTNATHDEVREG--LNQLNQIHEKAKNDVNSQTNQOVE 2843
Qy 1732 NAPLPLMKYGLFQGNPFKTDQWRLVFQNDNDQIAALRVQEQDR-----PEKSSDEKD 1786
Db 2844 NAE-----QNSLD-QINNFRPDFSKRNAVAEIVKAQOQNKIDEIEQEFSAQEE 2891
Qy 1787 KQWKIKFKVPIPEMFNS-----GNIRFVGVMQIOGPNLWLPVIN-----1827
Db 2892 KDNALQHLDEQVKEIININSINQANTDNEVDNAKTSGLANI-----TEYRPEYNNKKNAILKL 2947
Qy 1828 -----SSVTDYFRCGTGDSNDVAN--LNVAPMQVKT---IAFTNNAFNNVKFEFNISK 1875
Db 2948 YDVSDTQEAINGPYDATEDBELQEAANSKLNKILLDAKQIGLAHTNNEVDIYNE--VSQ 3005
Qy 1876 KI 1877
Db 3006 KM 3007

RESULT 12
US-08-755-587-44
; Sequence 44, Application US/08755587
; Patent No. 6045997
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Ashworth, Alan
; APPLICANT: Stratton, Michael R
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer
; TITLE OF INVENTION: susceptibility gene and uses thereof.
; NUMBER OF SEQUENCES: 222
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,587
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-755-587-44

Query Match 2.6%; Score 256; DB 3; Length 3418;
Best Local Similarity 19.2%; Pred. No. 5.3e-08;
Matches 365; Conservative 258; Mismatches 656; Indels 626; Gaps 92;

QY 85 EFTDVSFKDFLTNGHRTVLEIPKQVWISERSPE-----DDKERF 126
DB 1126 EFTQF-RKPSYILQ--KSTFEPENQMTILKTTSEECRDADLVIMNAPSIGQVDSKQF 1182
QY 127 RLGFHLKEKLEDGNGIAQSAFTYLLPLDMPKALGOYSYIVDKN----- 171
DB 1183 EGTVEIKR-----KFAGLLKNDCKNSAGS---YLTDENEVGFRGFSAGTKL 1227
QY 172 -----FNNLIHPLSNFSAQSIKPLALTRSS--DFIAKLQFNNODELWVYL 216
DB 1228 NVSTEALQKAVKLFSD--IENISEETSAEVHPISLSSCKHDSVSMFKIENHNDKTVE 1285
QY 217 EKFFDLEALKANIRLQTADFSPEKGNLVDPFVYSFTRNPQV----- 258
DB 1286 KNNKQILQNNIEMTT-----GTFVEBITENYKRENTEDNKYTAASRNSHNLFDG 1338
QY 259 -----KEWASDLNQDKTVRLYLRTFSPOAKTILADKYKDETRFLSSIDLK- 305
DB 1339 SDSSKNDTVCIHKDETOLLFTDQHNICLKSGQFMKEGNTQIKR-DLSDLTFLVAKAQE 1397
QY 306 -----ASNGTSLFA--NENDLKQDQLOVDLDVSDYFGQSEITITSNSQVKKPVPASERSLK 358
DB 1398 ACHGNTSNKEQLTATKEQNIKD-----FETSD-----TFQATAGKNIISVAKELFN 1444
QY 359 DRVKFKDQOKPRIEKESL-----YFYDALSPY-----SOL 389
DB 1445 KIVNF-PDQRPPELHNFSLNSLSHSDIRKKNMBDILSYEETD1VKHKILKESVPGVTGNQ 1503
QY 390 QELVSKENSIKDLVNAFL-----ARNLAFSLGKYNFLFD----- 423
DB 1504 VTQCGQPERDEKIKPTLLGFHTAGSKVKVIAKESLDKVNLFDEKEQGTSEITSFSHOW 1563
QY 424 -----DLASHLDYFVLVSKAKIKQSSITKK-----LFIELPKISLKSSILGQ 467
DB 1564 AKTKYREACKDLELACETIETIAAPCKEMQNSLNDKNLVSIETVPPKLLSDNLRCQ 1623
QY 468 EPNIKTLFEKVTFLKDNPRDVEIEKAFGLLYGVNEE-----LEQAKAQRASFEKS- 522
DB 1624 TENLKT--SKSIFLKVKVHENVEKETAKSPATCYTNQSPYSVVIENSALAFYTSCSRKTSV 1681
QY 523 -----KKGLKE--PSQOKEENSAINNQELEDNITERLPENSPIQOQENAGLG 572
DB 1682 SQTSLLEAKKWLREGIFDQOPERINTA-----DYGVNLYENNSNTIAEN- 1727
QY 573 ASPDKPYMIKDQVQRYYLAKSQIQLIKAKDYTKLAKLSNRHTYNI SLRLKEQLFDVN 632
DB 1728 ---DKNHL---SEKQDTYLSNSS-----MSNSYSYH-----SDEVYN-D 1759
QY 633 PRIPSSRDEKAFVLDKTEKNKYWQIYSSASPVFQNKWSLFGYRYLLGLDLPQTIHEL 692
DB 1760 SRLSKNKLDSG---IEPVLKVEDQKNTSFSKVISNVKDANAY-----PQTVNE- 1806
QY 693 VKLGQKAGLQFEGYENLPDFNLKDNIRIKTLPFSOKDNFKLSLLDNNYVDGRIKAP 752
DB 1807 -----DICVEEL--VTSSSPCKKNKNAIKLSISNSNF---EVGPP 1842
QY 753 EFGPLPLPKELRRNSNGSGSQNSNPSWQE:ISQFKD-----QNL:SNQDOLA 801
DB 1843 APRI-----ASGKIRLCS---HETIKVKVDIDSFSKVIFTKENNENKSKIC 1885
QY 802 QFSTKIW-----EKIIGENEFQDNRLQYKLLDKQESWINKTRDNLWYTLGD 851
DB 1886 Q--TKIMAGCYEALDDSEDTLHNSLNDDECSMHSHKVFADIQOSEELQHNQNM--SGLEK 1941
QY 852 KLVKPKGN-NLEAKFRQISNLQELLTAFYTSAAALSNNNNYQDSGAKST----- 899
DB 1942 VSKISPCDVSLFETSDICKCSIGKL-----HKSVSANTCGIFSTASGKSQVQVSDASLQNR 1997

RESULT 13

US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700
City: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328,254
FILING DATE: 24-OCT-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/141,239

FILING DATE: 22-OCT-1993

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-CJ 1191

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 2482 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-328-254-6

Query Match 2.6%; Score 255; DB 1; Length 2482;

Best Local Similarity 18.9%; Pred. No. 3.8e-08;

Matches 378; Conservative 333; Mismatches 764; Indels 524; Gaps 91;

QY 61 NPTSDYQSVKALLNGKTFDPKSSFTDFVSKDFL-TNNGRTVLEIPKQVWISBSP 119
DB 434 NRKNELEQKEAF-----AKEHQBFLLKLAFAERNQMLDELTVOQALRSEMTD 484
QY 120 EDDKERFRLGHLKBEKEDGNIQAQATKFIYLLPLDMPKAAALGOYIVDKNFNNLI--- 176
DB 485 NON-----NSKSEAGGLKQE-----IMTLKEQNK-----MQEVDNLLQEN 521
QY 177 -----IHLPSNFSQSIKPLALTRSSDFIAKLQ--FNNQDELWV-----YLEK 218
DB 522 EQLMKVMTKHCQNLSEPIRNSVKERESE-----RNQCNFKPQMDLEVKEISLDSYNAQ 577
QY 219 FFDLEALKAN-----IRLOTADFSFEKGNLVD-----PFV 248
DB 578 LVQLEAMLRNKLKQLQSEKEKECELOHELTQIRGDLTSNLDQMOSQISGLKQCEIDAE 637
QY 249 YSFIRNPQKQEWASDLNQDQTVRLYLRTFS-----PQAKTILKDYKYKDETFLSSID 303
DB 638 EKIYSGPHE-----LSTSQNDNAHLQCSLQTTWNKLNLEKEICEILQAEKVELTELNDNR 693
QY 304 LKASNGTSIFANE-----NDLKQDLVDLLDVSDFGGQ-SETITSNSQKVPV 351
DB 694 SECITATRKMAEEVGKLLNEVKILNDDSGLLHGLL--VEDIPGGEFGEPQNEQHPVSLAP 751
QY 352 ASERSLKDVRKFKDQKQKPR-----EKFSLEYDALSFSYQQLVSKPNSIKDLVNATL 407
DB 752 LDESNSYELHTLSDKEVQMHFAELQEKFLSLQSEHKILHDQHCQSSKMSSELQTTVDSLK 811
QY 408 ARNLRFSLGKYNF---LPDDLASHLDYFEL--VSKAKIKQSSITKKL-----FIEPL 454
DB 812 AENLVLTNLRFQGLVKEMQGLGEEGLVPSLSSSCVPDSSLSLGLDSSFYRALLEQT 871
QY 455 IKISLSSILGQEPN-----IKTLFEKVTFKLDFRDEVEIEKAPGLLYPGVNEEL 506
DB 872 GDMSSLNLEGAVSANQCSVDEVFCSSLQEENLTRK-----ETPSAPAKGV-----EEL 920
QY 507 EQARKAQRASFKEKSKGLKEFSQOKE--ENSKAINNQGLEBDDNITERLPENSPIQVQ 565

DB 921 ESLSCEVYVRSQLEKLEKESQGIKMKETQLEQLLSERQELDLCLRKQYVLSNE--OMQ 978
QY 566 QENAGLGASDPKPYMIKDVNQRYYL---AKSOIQELIKAKDYTKLAKL-----611
DB 979 OKLTSVTLMESEKLAAREKQTEQLSLELEVARLOQL-----DLSRSLGIDTEDIAQ 1034
QY 612 -----LSNRHTYINISIRLKE---OLFVDPNPRIPSSRDIEKAKFVLDTKTKNKYMQIYS 661
DB 1035 RNESCDSIKHETSETTERTPKDVHQCIDKAOODLNLDEK---ITETGALAPTGCSG 1091
QY 662 SASPFQNKWSLFGYYRYLLGLDPKQTIHELKVKQKAGLOFGYENL-PSDF--NUEDL 718
DB 1092 EQSPDTNYEPP-----GEDKTQGSSECI-----SELSFGPNALVPMDFLGNQEDI 1137
QY 719 KNTRIKTPLFSQKONFKLSLLDFNNYVDGEIKAPEGLPLPLPKELRRNSSSGSQNSN 778
DB 1138 HNLQLRV---KETSNNENLRLLHVIEDRDRK-----1165
QY 779 SPWEOETISQFQKONLSNOQLAQFSTKI---WEKIIGD---ENEFQDNRRLOQ---826
DB 1166 -----ESLLNEMKELDSKLHLQEVQLMTKIEACIELEKIVGELKKENS-DLSEKLEYSCD 1220
QY 827 --KILK-----DLOESWINKTRDNLVYVGLDKLVKPKNNLEAKPRQISNLOEL 874
DB 1221 HQELLQRVETSEGLNSDLEMHADKSSRED-----IGDNV-AKVNDWSKBRFLDVEN--EL 1272
QY 875 LTAFYTSAAALSNMNNYQDSGAKSTIIFEBIAELDPKVKKEKGVADVVQLKHFHYAIGFDDN 934
DB 1273 SRIRSEKASIEHEALYLE-----ADLEVVOQTEKLCLE-----KDN 1307
QY 935 AGFNQOEVRSSRTIYKTSKSKL--EADT-----IDQLNQAVKNAPLAGLSFYLD 985
DB 1308 ENK--QKVIIVCLSEELSVTSERNQALGELDTWSKTTALDOLSEKMKETQELSEHSE 1365
QY 986 TERFGVFKLATSIAVQHKQKKT-LPKKLNNDGYTLIHDKLKKPVTPQISSSEKPD---1041
DB 1366 C-----LHCIQVAEAEVKEKTELLQTLSSDVELLKQKTH---LQEKLSLEKDSQA 1414
QY 1042 -----WFEGLKQNGSQNVV---STFGSIIESPYFSTNFOE--DADLDQDG-----1084
DB 1415 LSLTKCLENQIAQANKEKELLVKESLOARSESDEYKLVNSKALAEALVKEGFAIR 1474
QY 1085 -----QDSSROGNNSLDNQEAGL-LKQKLAITLLGNQFIQYVQONDKEIEFEIINVKVSEL 1139
DB 1475 LSSTQEEVHQLRGIEKLRVRIEADKQLHIAEKLEREREND-SLKDKVENLERELQM 1533
QY 1140 SFRVEFKLAKTLEDNGKTIIRVLSDETWSLI-VNTTIETKTPMSAVPEVPDTKWVQYDPR 1199
DB 1534 S-----EENQELV-ILDAENSKAEVETLTKQTEEMARSLKVPFELDLVTLRSEKE 1581
QY 1200 PLAAKTFVLKFKDQIPVDGSGNIS--DKMLASIPVLIHQOMLRSLPVVKTIRRELGLKTE 1257
DB 1582 NLTRQIQ-----EKGQISLSELDKLSLSPKSLLE-----KEQAEIQIKBE 1621
QY 1258 QQQQQQQQQQQQQQPKKAVRK---EDEL-----ETYNP--KDEFNINPLTKAHLRLTSLN 1307
DB 1622 SKTAVEMLONQKELNEVAALCGDQEIWKATEQSLDPPPIEEHQLRNRSIEK-----LRA 1676
QY 1308 LVNNDPNYKIBDLKVIKNEAGDHQALFSLRANNIKRLMNTPIITADYNPPFYFNEDWRSI 1367
DB 1677 RLEADEKQKQCVLQOLK-ESBHADLLKGRVENLERELEIART-----NOEHAAL 1725
QY 1368 KYLNKNGVSSHQOQAAGNGQ--SGL-IQRLNKNIKETPTTALIALKDR--NNTNLGNY 1424
DB 1726 EAE-NSKEVETLAKITEGMTQSURGLDELVDVVTTRSEKENLTNELQKEQERISELEIINS 1784
QY 1425 S-DKTIIMTKPKYLVERSIGVFWSTGLDGYIGSEQTKDGTSSSSQKQKQDPQDFIQAIGLKN 1483
DB 1785 SFENILOKEQEKV-----QMKESSTAMEML---QTKLKNELNVR 1822
QY 1484 TEYHG-----KLGLSIRIFDPGNELAKIKDANSKYGEKLLKSYDLFKNYLN-----1530

```
Db 1823 AALHNDQEAACKAEQNLSSQV-----ECLELEKAQLQGLDDEAKNNYIVLQSSVNGLIQE 1877
Qy 1531 -----EYEKKSPKIAKGWTHNIHPDQKEYPNPQKLPENYLNVLNQPKVTLNYSDDFI 1584
Db 1878 VEDGKQKLEKDEISRLKNQIQDQEQVSKLSQVEGHQL-----WK-----EQNLLE 1926
Qy 1585 TNLFPVEPESDRSGTKLKQVIOKQ-----VNNYADWGSAYLTFWYDKNIITNQ 1634
Db 1927 RNLTVLEL-----QKIVLQSKNASLQDTLLEVQLSSYKNLELELTQMDKMSFVEK 1978
Qy 1635 PNVTIA-----NIADVFIKDVKELEDNTKLJAPNITQWMPNIGSGKFKYKPTVF 1684
Db 1979 VNKWTAKETELQREHMAQTAELQEEISGEKNRLAGELQLLLEIKSKQGLKELTL- 2037
Qy 1685 FGNWENENSMNS-----QAQPTTWBKIREGFALQALK-SSPDQKTRTFVLTNAPLEP 1737
Db 2038 -----ENSELKSLDCMHKQVKEGKVEEAEYQLRLHAEAKKHQALLDNTNQYEV 2091
Qy 1738 WKYGPGLGFQNGPFKTDWRFLVFQNDNDQIALURVQEDRPEKSSDDKQKWKIKFVVI 1797
Db 2092 -----EIQTYREKLTSKEECLSSQKL-EIDLKSSKEELNNS--LKATTQI 2134
Qy 1798 PERMFNS--GNIRFVGVMQ 1814
Db 2135 LEEELKTKMDNLKYNVQLK 2153

RESULT 14
US-08-639-501-2
; Sequence 2, Application US/08639501
; Patent No. 5837492
; GENERAL INFORMATION:
; APPLICANT: Tavtigan, Sean V.
; APPLICANT: Kamb, Alexander
; APPLICANT: Simard, Jacques
; APPLICANT: Couch, Feigus
; APPLICANT: Rommens, Johanna
; APPLICANT: Weber, Barbara
; TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti
; STREET: 1201 New York Avenue N.W., Suite 1001
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/639,501
; FILING DATE: 29-APR-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/585,391
; FILING DATE: 11-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/576,559
; FILING DATE: 21-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/575,359
; FILING DATE: 20-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/573,779
; FILING DATE: 18-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-116802-04
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-639-501-2

Query Match      2.6%; Score 253; DB 2; Length 3418;
Best Local Similarity 19.1%; Pred. No. 8.5e-08;
Matches 364; Conservative 258; Mismatches 657; Indels 626; Gaps 91;

Qy 85 EPTDVSFVDFLTNNGRVTLEIPKPKYQVVISFSP-----DDKERF 126
Db 1126 EFTQF-RKPSYILQ--KSTFEVPENQMTILKTTSECRDADLHVIMNAPSIGQVSSKQF 1182
Qy 127 RLGFHLKEKLEDNIAQSATKFIYLLPLDMPKAALQGVSYIVDKN----- 171
Db 1183 EGTVEIKR-----KPAGLLKNDCKNSAGS---YLTDENEVGPRGFYSAHGTKL 1227
Qy 172 -----FNNLIHPLSNFSAOSIKPLALTRSS--DFIAKLQNFNQDELWVYL 216
Db 1228 NVSTEALQKAVKLFSD--IENISEETSAEVHPISLSSSKCHDSVSMFKIENHNOKTVSE 1285
Qy 217 EKFFDLEALKANIRLQTADFSPEKGNLVDPPFYYSFIRNPQO----- 258
Db 1286 KNNKQQLILQNNIEMTT-----GTFVEEIEFENYKRNTEENEDNKYTAASRNSHNLDPG 1338
Qy 259 -----KEWASDLNQDKTVRYLYLRTFSPQAKTILKYKYKDETFLSSIDLK- 305
Db 1339 SDSKNDYTCIHKDETDLLFTDQHNICLKLSQCFMKEGNTQIKE-DLSDLTFLVAKAQE 1397
Qy 306 -----ASNGTSLFA--NENDLKQDLVDLLDVSDFGQSEFITSNSQVKKVPASERSLK 358
Db 1398 ACHGNTSNKEQLTATKTEQNIKD-----FHTSD-----TFFQTASGKNISVAKESFN 1444
Qy 359 DRVFKPKDQOKPRIEKFSL-----YHVDALSFY-----SOL 389
Db 1445 KIVNF-FDQKPEELHNFNSLHSDIRKKNWDILSYBETDIVKHILKESVPGTGNQL 1503
Qy 390 QELVSKPNSIKDLVNATL-----ARNLRFSLGKYGNFLFD----- 423
Db 1504 VTFQQPERDEKIKBPTLLGFHTASGKVKIAKESLDKVNLFDEKEQGTSTITSFSHOW 1563
Qy 424 -----DLASHLDYFVLSKAKIKQSSITKK--LFIELPIKISLSSITLGDQ 467
Db 1564 AKTLKYREACKDLELACETIEITAAPCKEMQNSLNNDKNLVSIVTVPVKLLSNLCRQ 1623
Qy 468 EPNIKTLFEKEVTFKLDNFRDVEIEKAFGLAYPGVNEE-----LEQARKAQRASFKEKS- 522
Db 1624 TENLAT--SKSIFLKVKHVENVEKETAKSPATCYTNQSPYSVIENSALAFYTSCSRKTSV 1681
Qy 523 -----KGLKE--PSQOKEENSKAINNQGLEEDDNITERLPENSPFIQOQENAGLG 572
Db 1682 SQTSLLEAKKWLREGIFDQGPERRINTA-----DYVGNLYENNSNSTIAEN----- 1727
Qy 573 ASPDKPYMIKDVONORYYLAQSIQELIKAKDYTKLAKLLSNRHTYNIISRLKEOLFVN 632
Db 1728 ---DKNHL---SEKQDTYLSNSS-----MSNSYSYH-----SDEVYNS 1760
Qy 633 PRIPSSRDIEKAKFVLDKTEKNKYMQIYSSASPVFQNKWSLFGYYRYLLGLDPKQTHIEL 692
Db 1761 GYLSKNKLDSGIEPVLKNVEDQK---NTFSKVISNVKDANAY-----PQTVNE- 1806
Qy 693 VKLGQKAGLQFEGYENLPSDFNLELDKNIRIKTPLFSQKDNFKLSLLDPNNYDGEIKAP 752
Db 1807 -----DICVEEL--VTSSSPCKNNAIKLSISNNP---EVGPP 1842
Qy 753 EFGPLFLPKELRRNSNSGGSQNSPWEQEIISQFKD-----QNLNQDQLA 801
```

1843 AFRI-----ASGKIVCVS---HETIKKVKIDFTDSFSKVIKENNENSKIC 1885
802 QFSTKIW-----EKIIGDNEFDQNNRLQYKLLKLOESWINKTRDNLWTVLGD 851
1886 Q--TKIMAGCYEALDDSDGILHNSLDNDECSTHSHKVFADIQSEILQHNQM--SGLEK 1941
852 KLVKPKN-NLEAFPRQISNLOELLTAFYTSAAALNNWNYQDSGAKST-----899
1942 VSKISPCDVSLETSDICKCSIGKL-----HKSVSANTCGIFSTAGSKSVQVSDASLQAR 1997
900 IIFEIABLPKVKKEV-----GADVYQLKFHYAIGFDD---NAGKFNQEVIRSSRTI 950
1998 QVFSEIEDSTQVFSKVLFKSNEHSDQLTRENTAIRTPHELIQKGFYSNVNVSASFSG 2057
951 YLKTSGSKLEADTIDQINQAVKNAPIGLQSF-----YLDTERFGVFKLATSLA 1000
2058 FSTASGK--QVSILESLSHKVKG--LEBFDLIRTEHLSHYSPTRQNV-----2102
1001 VOHKQKEKTLPKLNNDDGYTLIHDKLPKVPVPOISSPEKDFEGKLNQSGSQNVNVT 1060
2103 -----SKILPRVDKRNPEHCNVSEMEKTCSEKFKLS-----NNLVGGSSENK---2146
1061 FGSIISEPYFTNFQEDADLDQDGDQDSDRQGNNSLDNQEAGLLKQKLAAILGNQFIQYQ 1120
2147 --HSIKVSYLS-QFOOD-----KQQLVLGTVKSVLNIHVLGKEQ 2184
1121 QNDKEIEPEIINVEKVSLSFRVEPKLAKT-----LEDNGKT-IR 1159
2185 ASPKNVMEIGTETFSVPVKTNIETVCSYKDSSENYFETEAVEIAKAFMEDDELTDK 2244
1160 VLSDETMEL-----IVNTTIEK---TPMSAVEPEVDTKWVEQYDPRTPLAATK 1206
2245 LPSHATHSLFTCPENEEVMSNRIGKRGKRGFLILVGEPSIKRNLNLEFDRIIENQEK-S- 2303
1207 FVLKPKDQIPVDGSGNIGDKWLASPLVIHQMLRLSPWV---KTIRELGLKTEQOQQQ 1262
2304 --LKASKSTP---DGTIKDR-----RLFWMH--VSLEPITCVPPRTTKE-----2340
1263 QQQQQQQQPKKAVRKEBELETYNPKDEFNLPITKAH---RLTL-----SNL-VNNDPN 1314
2341 --RQBIQNPFTA-----PGQEF-----LSKSHLYEHLTLEKSSNLAWSGHFF 2382
1315 YKIEDLVKIKNEAGDHLAFSLRANNIKELMNTPTITPADYNPFFYNNEDMSIDKYLNN- 1373
2383 QVVS-----ATRNKMRHLITTRPTKVFVPPFKTSHFHRVEQCVRN 2426
1374 -----KGVSSHQQAGGQSGSLTORLNNKIKPE---TFTPA-----LIALKDR 1416
2427 NLEBRQKQNDIGHGSDSKKNKINDNEIHFQKNNSNQAAVFTFKCEEPLDLI-----2481
1417 NNTNLSNYSK-KIIMIKPK-----YLVERSIGVPMSTGLDGYIGSEQTKDGT 1463
2482 --TSLQARDIQDMRIKKQRQVFPQPSGLYLAKTST-LP-RISLKAAGVQVPSACSH 2537
1464 SSSQKQGFQDQIALGLKQTE-----YHCK-----LGLSIRIFDPCNELAKIKDASN 1511
2538 KOLYTVGVSKHCIK-INSKNAESFQHTEDYFKESLWTKGIQIADGG---WLIPSDNG 2593
1512 KKEEKLKLSYDLFKNYLNEYKKSPK-IAGQWNIHPDOKEYPNPQKL-----PEN 1563
2594 KAGKEB-----FYALCOTFGVDPKLSRIWVNH-----YRWIWLKLAWECAFPKE 2641
1564 YLNLVLNQPKWTVLNSDFITNLFVEPEGSDRGSGTKLKQVIOKQVNNNYADWGSAYLT 1623
2642 PANRCLS-PERVLQKLYRYDTEI-----DRSRSRAIKKIMER-----DDTAATKL 2686
1624 FWDKNNIINQPNVITANTADVFIDKVELENTKLIAPNITQW 1668
2687 VLVSDIIS-----LSANISSETSSNKTSS-ADTQKVAIIELTDGW 2725

RESULT 15
US-09-044-946-2

; Sequence 2, Application US/09044946
; Patent No. 6033857
; GENERAL INFORMATION:
; APPLICANT: Tavtighian, Sean V.
; APPLICANT: Kamb, Alexander
; APPLICANT: Simard, Jacques
; APPLICANT: Couch, Fergus
; APPLICANT: Rommens, Johanna
; APPLICANT: Weber, Barbara
; TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti
; STREET: 1201 New York Avenue N.W., Suite 1001
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/044,946
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/639,501
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/576,559
; FILING DATE: 21-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/575,359
; FILING DATE: 20-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/573,779
; FILING DATE: 18-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-116802-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-044-946-2

Query Match 2.6%; Score 253; DB 3; Length 3418;
Best Local Similarity 19.1%; Pred. No. b 5e-08;
Matches 364; Conservative 258; Mismatches 657; Indels 626; Gaps 91;
QY 85 EFTDFVSKFDFLTNNGRTVLEIPKQYQVVIIEFSPSE-----DDKERF 126
DB 1126 EFTQF-RKPSYILQ--KSTFEVPEPQMTILKTTSECRDADLHVIMNAPSIGQVDSKQF 1182
QY 127 RLGFHLKEKLEDGNIAGSATFYILLPLDMPKALQGYSIYVDKN-----171
DB 1183 EGTVEIKR-----KFAGLLKNDCKNSAGS---YLTDENEVGFRGFYSAHGTKL 1227
QY 172 -----FNNLIHPLSNFSAQSIKPLALTRSS---DFTAKLNQFNQDELWYVL 216
DB 1228 NVSTALQKAVKLFSD--IENISETSAAVPIISLSSKCHDSVVMFKIENHNDKTYS 1285
QY 217 EKFFDLEALKANIRLQATDFSPFEKGNLVDPFVYFIRNPQNO-----258

Db 1286 KNNKCOLLQNNIEMTT-----GTFVEBITENYKNTNEDNKNYTAASRNHNLLEFDG 1338
Qy 259 -----REWASDLNQDQKTVRLYLURTFSPQAKTILKDYKYKDBTFLSSIDLK- 305
Db 1339 SDSSKNDTVCIHKDETDLLFTDQHNIKLSGGQFMKEGNTQIKE-DLSDLITFLEVAKAQE 1397
Qy 306 -----ASNGTSLFA--NENDLKOQLDVLDDVDSDYFGQSETITSNSQKVPVPAERSLK 358
Db 1398 ACHGNTSNKESQLTATKTEQNIKD-----FETSDD-----TPFQASGKNISVAKESFN 1444
Qy 359 DRYKFKKDOQKPIEKPSL-----YEXDALSFY-----SOL 389
Db 1445 KYNF-PDQPEBLHNFSLNSLHSDIRKKNMILSYEETDVIKHKILKESVPVGTGNQL 1503
Qy 390 QELVSPNSIKDLVNATL-----ARNLRFSLGKYNFLFD-----423
Db 1504 VTFQGPDERDEKIEPTLLGFHTASGKKVIKAKESLDKVNLFDEKEQGTSEITTSFSHOW 1563
Qy 424 -----DLASHLDYFVLVSKAKIKQSSITKK-----LFIEPLIKLSKSSILGDO 467
Db 1564 AKTLKYREACKOLELACETIETIAPACKEMQNSLNDKMLVSIETVPPKLLSDNLCRQ 1623
Qy 468 EPNIKTLFEKEVTFKLDNFRDVELEKAFGLLYPGVNEE-----LEQARKAQRASFEKEKS- 522
Db 1624 TENIKT--SKSIFLKVKHENVKEKTAQSPATCYTNQSPYSVIENGALAPYTCSCRKTSV 1681
Qy 523 -----KKGLKE--FSQOKEENSKAINNQEGLLEDDNITERLPENSPITQYOENAGLG 572
Db 1682 SQTSLLEAKKWLREGFDGQPERINTA-----DYVGNLYENNSNSTIAEN--- 1727
Qy 573 ASDPKPMIKVDONQRYYLAKSOIQELIKAKOYTKLAKLSNRHTYNIURLKELQFDPVN 632
Db 1728 ---DKNHL---SEKQDTYLSNSS-----MSNSYSYH-----SDEVYNS 1760
Qy 633 PRIPSSRDIEKAFVLDDKTEKNYQWQIYSSASPVFQKNKSLFCGYRYLLGLDPKQTIHEL 692
Db 1761 GYLSKNKLDGIEPVLKNVEDQK-----NTSFSKVISNVKDANAY-----PQVNE- 1806
Qy 693 VKLGQKAGLOFEGYENLPDNLKNIIRIKTFLFSQKDNFKLSLLDFNNYDGBIKAP 752
Db 1807 -----DICVEEL--VTSSSPCKNKNAAIKLSISNSNF---EVGPP 1842
Qy 753 EFGLPFLPKELNRNSNSGSGSONSPWQEBIISQFKD-----QNLNSQDOLA 801
Db 1843 AFRI-----ASGKIVCVS---HETIKKVKDIFTDSFSKVIKENNENSKIC 1885
Qy 802 QFSTKIW-----EKIIGDENBFDONRLQYKLLKDQESWINKTRDNLWYTVLGD 851
Db 1886 Q--TKIWAGCYEALDDSEDILHNSLDNDECSHSHKVFADIQSEELQHNQNM--SGLEK 1941
Qy 852 KLKVKPKN-NLEAKFRQISNLQELLTAFYTSAAALSNNWNYQDSGAKST-----899
Db 1942 VSKISPCDVSELTSDICKCSIGKL---HKSVSANTCGIFSTASGSKVQVSDASLQNR 1997
Qy 900 IIFEETAELDPKVEKY-----GADVYQLKFHYAIGFDD---NACKFNOEVRSSRRTI 950
Db 1998 QVFSEIEDSTQVFSKVLKFSNHSQDLTREENTAIRTPHEHLISQKGFSNVNVNSAFSG 2057
Qy 951 YLKTSGSKLEADTIQOLNOAVKNAPLGLQSF-----YLDTERFCVFOKLATSLA 1000
Db 2058 FSTASGK---QVSILESSLHKVGV---LEEPDLIRTEHLSHYSPTRQNV-----2102
Qy 1001 VOHQKEKTLPKKLNDNGYTLIHDKLKPKVIPQISSPEKDMFEGKLNQNGSQNVNST 1060
Db 2103 -----SKILPRDKRNPCHVNSEMEKTSCKEFLS-----NNLNVEGGSSENN--- 2146
Qy 1061 FGSIIESPSTFQEDADLDQGDQDSRQGNNSLNOEAGLLKQKLAILLGNQFTQYYQ 1120
Db 2147 -HSIKVSPYLS-QFQD-----KOQLVLGTQKVSLENIHVILGREQ 2184
Qy 1121 QNDKETEFELINVEKVSLSFRVEFKLAKT-----LEDNGKT-IR 1159
Db 2185 ASPKNVQMEIGKTETTSDDVPVKNIIEVCSTYKDSENYFTEAVEATAKAFMEDDELTDK 2244

Qy 1160 VLSDETMSL-----IVNTTIEK---TPEMSAPVEFDTKWVEQVDPRTPLAAKTK 1206
Db 2245 LPSHATHSLFTCPENEBMVLNSRIGKRGRGEPILLVGEPSIKRNLNNEFDRIIENQEK- 2303
Qy 1207 FVLKFKDQIPVDGSGNISDKWLASIPLVTHQOMLRLSPVU---KTIRELGLKTEQOQQO 1262
Db 2304 --LKASKSPF--DGTIKDR---RLFMHH--VSLEPITCVPPFRITKE-----2340
Qy 1263 QOQOQOQOQKXAVRKEEBELETYNPKDEFNILNPLTKAH---RLTL---SNL-VNNDPN 1314
Db 2341 --ROELQNFNFA-----PQOEF-----LSKSHLYEHLTLEKSSNLAVSGHPF 2382
Qy 1315 YKIEDLKVIKNBAGDHLAFSLRANNIKRLMNTPTTFADYNPPFFYNNEDWRSIDKYLNN- 1373
Db 2383 YQVS-----ATRNEKMRHLITTRPTKVFVPPFKTSHFHRVEQCVRNI 2426
Qy 1374 -----KGNVSHQOQAAGGQSGSLIQRLNKNIKPE---TFTPA-----LIALKOR 1416
Db 2427 NLEENRQKQNDHGSDDSKNKINDNEIHOFKNQNSNQAAVFTTKCEBEPLDLI-----2481
Qy 1417 NNTNLSNYS-D-KIIMIKPK-----YLVERSIGVPWSTGLDGVIGSEQTKDGTS 1463
Db 2482 --TSLQNRARDIQDMRIKKQQRQVFPQPSLYLAKTST-LP-RISLKAAGVQGPACSH 2537
Qy 1464 SSSQKGFDPDQFIQALGLKQTE-----YHGK-----LGLSIRIPDPGNEIAKIKDASN 1511
Db 2538 KQLYTVGSKHKCIK-INSKNABSFQPHTRDYFKESLWTKGQIQLADGG---WLIPSDNG 2593
Qy 1512 KGEELKLSYDLFKNYLNEYEKKSPK-TAKGWTNIHPDQKEVPNPQKL-----PEN 1563
Db 2594 KAGKEB-----FYALCOTPGVDPKLSIRIWWYNH-----YRWIWLKLAAMECAPPK 2641
Qy 1564 YLNLVLNPKWVTLNYSDFITNLFVEPEGSRGSGTKLKQVIQKOVNNNYADWGSAYLT 1623
Db 2642 FANRCLS-PERVLQLKYRYDTEI-----DSSRSIAKKIMER-----DDTAATKL 2686
Qy 1624 FWYDKNIINQPNVITANITADVPFKDKBELBNTKLIAPNITQMW 1668
Db 2687 VLCVSDIIS-----LSANISSETSSNKTSS-ADTQKVAILIETDGM 2725

RESULT 16

US-09-044-908-2
; Sequence 2, Application US/09044908
; Patent No. 6124104
; GENERAL INFORMATION:
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Kamb, Alexander
; APPLICANT: Simard, Jacques
; APPLICANT: Couch, Fergus
; APPLICANT: Rommens, Johanna
; APPLICANT: Weber, Barbara
; TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti
; STREET: 1201 New York Avenue N.W., Suite 1001
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/044,908
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:


```
Db 1758 SDSKFRKYSQKVAEQAKIOQKVNMIQKIKTNKALNSAQRAQLQBELKQAKLDL 1817
Qy 1493 SIRIFDPGNEL-----AKIKDASNK-----KGEELKLLSYDL 1524
Db 1818 -ISVQDVRELQKLVQSKVDTELKSIKSSSKTQGIKVDNKSIMTEDEDDKVKYYSK 1876
Qy 1525 PKNYLNEYKSPKIAKGNTHHPDQKBYKPNPNQKLPENYLNVLVNPWK-----VTL 1577
Db 1877 QIKLIQOQKBEAKYIKQLEBQKKAAGFPDIOQITEEMQN-----WKDKQKDFNLEL 1930
Qy 1578 YNSDDFTNLRF-----VEPEGSDRGSGTKLKQV----- 1605
Db 1931 YNTKSGINDIYKSLADEVSVIYKEMYEKMRDIELEAHQKATQDLIDEIDKTDDEAKFQKE 1990
Qy 1606 -----IQK-----QVNNYADWGSAYLTFWYDKNIITNQPNVITANI-----ADVFID 1649
Db 1991 LKERQDSIQKLTQDQINQYSLD-----DSEFGSKVKVELTEQLQKEQLDLDLDFLKD 2040
Qy 1650 -----VKEEDNTKLIAPNITQWPNISGSKEKF-----YKPTVFEQNW 1688
Db 2041 RESNKRKEALQDLEKDEBESINNKYDNLVNDERAFKLEDKIMNGKITDIKQJNEFSKF 2100
Qy 1689 ENEN-SSMNSQAQPTTWKIREGALQALKSSFDQKTRTFVLTNAPLWKYGLGFQON 1747
Db 2101 INTNMESIGKISNNLIDKJKE--ASNALNTAVKGNITGKVKVSFAS---GGYGTGLGA 2155
Qy 1748 GPNFKTQDWRVLFQND-----NOTAALRVQEQDRPEKSSBDKQKQW---IKPKVWIPE 1799
Db 2156 GKLAFLHDKELILNKTDANILDTVKAVRETAVD-----DSPKQGVGLADLIKK 2206
Qy 1800 EMFSGNTRFVGWMOIQGNPTLWLPVINSVVIYDFYRGTGSDNDVANLWAPQVKTIAP 1859
Db 2207 -----GITSIPS-----LVBNVQNSML-----TNSLPNLKKEIPEKTIAS 2243
Qy 1860 TNAENNVKERNISKKI 1877
Db 2244 SGDKTINLTNTHIDKLI 2261

RESULT 18
US-09-932-183A-2
; Sequence 2, Application US/09932183A
; Patent No. 6833265
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394C1-US
; CURRENT APPLICATION NUMBER: US/09/932,183A
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-932-183A-2

Query Match 2.6%; Score 252; DB 4; Length 2285;
Best Local Similarity 18.1%; Pred. No. 5.5e-08;
Matches 395; Conservative 339; Mismatches 838; Indels 606; Gaps 94;

Qy 60 ANFTSDYQSVKALLNGKTFDPKSSEFTDFVS---KPDFLTNGRTVLEIPK-KYQVVIS 115
Db 330 AELTQTFKSMSTYLSGSLFYGAISGLKEMVQAIEIDTMTNIRRVNNEPDYKYNELL- 388
Qy 116 EFSPEDDKFRFLGHLKLEKLED-----GNIAQSATKFIYLLPLDMPKALGQYSYVDK 170
```

```
Db 389 -----QESIDLGDTLNSNKITDILQMTGDFGRMGFDESELSTLTTKTAQVLQNVSDLTDP 441
Qy 171 NFNNLIHPLSNFSAOSIKPLALTRSSDIFIAKLQNFQNDQELWVYLEKFPDL--EALKAN 228
Db 442 DTVNTLTAAMLNPN-----IAANDSISADKLNVEYDNN-----YAVTTLDLANSIRKAG 490
Qy 229 IRLQT-----ADFSPEKGNLVD--PFVYSFIRNPQNKQEWASDLNQDQK 270
Db 491 STATTGVELNDLIGYTTAISTTRESGNIVGNSLTIIFARIGNQSSIKALEQIGISVK 550
Qy 271 TVRLYLRTFESPOAKTILKDYKYKDETFIASSIDLKASNGTFS---LFAENNDLKQDLVDL 327
Db 551 TAGGEAKS-----ASDLISEVAGKWDLTSDAQKNTSIGVAGIYQLSRFAMNMFISI-- 603
Qy 328 LDVSDYFGGQSEIITSNSQVKPVPASERSLKDRVKFKKQDOOKPRIEKSLEYVEDALSFYS 387
Db 604 -----AQNAAKTAANSTGSAWSEQQKYADSLQARVKNLQNFNTEFAAASDA--FIS 653
Qy 388 QLQELVSKPNSIKDVLNATLARNLRFSLGKYNELFDDLASHLDYFYFLVSKAKIQSSITK 447
Db 654 --DGLIEFTQAAGSLNASTG-----VIKSVGFLPPLAA-----VSTATLLSKNTR 699
Qy 448 KLFIELPIKISLSKSIILGDQEPNIKTL-----FKKVTFFKLDNFRDVEIEKAF-----GLL 498
Db 700 TL-----ASSLILGTRAMQOETLATAGLEAGMT-----RAAVASRVLTALRGLL 744
Qy 499 YP-----GVNEELQARQAQASFEKSKKGLKFEFSQOKEENSKAINNQEGLEDDNITE 554
Db 745 VSTLVGGAFAALGAWLESLSF--AEAKKAKDDFEQSQQTNVEAITNK-----DSTDK 797
Qy 555 RLPENSPIQYOENAGIGASPDKPY-----MIKDQONQRYLAKSOIOLIK 601
Db 798 LIQYKELQVKKRSRSLTSDEBQEYLQVTOQLAQTFFALVKGYSQGNAILKTN--KELEK 856
Qy 602 AKDYTKLAKLLSNRHTYNIISLRKQLFDVNPRIIPSRD--IEKAFVLDKTKRKYWOIY 660
Db 857 AIENTKEYLAKKQETRDSA---KKTPEDAKEIKSKDELKQYKQIADYNDKGR----- 908
Qy 661 SSASPVFQKWSLFG--YRYLLGLD--PKQIHELKVLGQKAGLQFEGYENLPSPFNLEDL 718
Db 909 -----PKWDLTADDDYKYAADKAKQSM-----LKAQSDIES-- 940
Qy 719 KNIRIKTLPFSQKDNFKLSLLDFNNYDGEIKAPEGLPL---FLPKELRRNSNSGSGQ 775
Db 941 GNAKVKDSVLIANAY--SSIDISNTLTKTSIDVVVKLNKLDLDPELEKFSLSGLKLQ 998
Qy 776 NSNSPWEQEIISQPKDQNLNSQDQLAQFSTKIWEKIIIGDENEFQNNRLOKYLKLDQES 835
Db 999 EK-----NQKALDSGDEKAFDNNAKKDLQSLLETYSKSDSIDVFKMSFDKAKNIKDGDKS 1054
Qy 836 WI-----NKTRNLTYWTYLGDKLK-----VKPNKLEAKFRQIS----- 869
Db 1055 LSSVKSEVGDIGETLABAGNEAED-----FGKKLKEALDANSVDDIKAAIKEMSDAMQF 1108
Qy 870 -NLOELLTAFYTSAAALSNNWNYODSGAKSTIIFEEIAE-----LDPKVKKEVG 917
Db 1109 DSVQDVLN-----GDFNNTKQOVAPLNDLLEKWAEGKSISANEANTLIQKDE--- 1157
Qy 918 ADVYQLKPFHYAIGFDDNAGFNQ--EVIRS-----SSRTIYLTSGSK--LE 961
Db 1158 -----LAQAI SIENGWVKINRDEVIKQKRVKLDAYNDVMYTNKMLKMTVEVNAIKTLN 1210
Qy 962 ADT---IDOLNOAVKNAPLGOSFYLDTERFGVOKLATSLAVQHKQKTKLPKLNNDGY 1019
Db 1211 ADTLRIDSLKLRKERKLDMS-----EAELSDLVKSINNADAKKELKLEKMLQPGGY 1266
Qy 1020 TLJH-----DKLKKPVIQISSSPKDWFEGLKNGQSGQNVNVSTFGSIIIESPVSTNFQ 1075
Db 1267 SNSQIEMQSVKSALESYISASEAT-----STQENWKO---ALVEA---GTSLE 1310
Qy 1076 EDAOLDQGDQDSRQGNNSLDNQBAGLLKOKLAILLGNQFTQYYOQNDKEIEFEIINVEK 1135
```

1311	NWTDQKQANBETKYSMVVDYKYEALKEVNAEDKYNQVNDYPKYYSQKYRDAIKKEIK	1370
1136	VSELSFRVFKLAKTLEDNGKTRIVLSDETMSLIVNTTIEKTPEMSAPVPEFDTKWVEQY	1195
1371	ALOQKKLQMEOAKLLDQIKSGNI---TQYGIVTSTSSGGTPSSTGSGYSKGYSYI	1426
1196	DPRTPLAAKTFVLKFKDQTPVDGSGNISDKWLASILPLVTHQO---	1240
1427	N-----SAASKY-----NVDP-----ALIAAVIQBESGFNAKARSVGVGAMGLM	1464
1241	RLSPVWKTIRELGLKTEBOQOQOQ---QOQOQOQOPKAVRKEBELETYPN-----	1287
1465	QLMPA---TAKSLGVNAYDPQNVMGTKYLAQOLEKFGNVKALAYNAGPGNVIKYG	1522
1288	-----KDFNILNPLTKAHLRTJLSNLVNDPNYKIED-----LKVKNBAG-----DHQ-	1331
1523	GIPPFKETQNVKKIMANYSKSLSSASTIASYYTNNNSAFRVSSKYGQOSSGLRSSPHKG	1582
1332	LAFSLRANN-IKRLMNTPIITFADYNP---FFYYNEDWRSIDKVLN---NKGNVSSHQOQA	1384
1583	TDFAAKAGTAKISLQSGKVQIAGYSKTAGNWNVITKDDGTIVAKYMHMLNTPSVKAGOSVK	1642
1385	AG-----GNQGSGL-----IQLRNKNIKPEFTFPAL-IALKDRNNT-----	1419
1643	AGQTIGKVGSTGNSTGNHLHLQIEQNGKTI DPEKYMOGIGTISDSASQAEARQOGIAQA	1702
1420	-----NLSNVSDKIIMIKPKYL-----VERSIGVPWSTGLDGYIGSEQ-----	1457
1703	KSDILLSLQGDITSSVNDQIQELQYELVQSKLDEFKIRIG-----DFDVRIAKDESMA	1757
1458	-----TKDGTSSSSQOKGFQDDFI-----QALGLKNTVEHGKLG	1492
1758	SDSKFPKYTSDDKKA VABQAKIQOKVNIQKIEKTNKALNSAQRAQLQEBELKQAKLDL	1817
1493	SIRIFDPGNEL-----AKIKOASNK-----KGEELKLSYDL	1524
1818	-ISVODQVRELQKQLVQSKVDETLKSIEKSSSKTQGIKDVDNKISMTEDDEKVKYYSK	1876
1525	FKNYLNEYEKSPKIAKGWNIHPDQKEYPNPNQKLPENVLNLVLNQPWK-----	1577
1877	QIKLIQOQOQKBAKYIKQLBEQKAAKGFDDIQOQITEMQN-----WKDKQKDFNLEL	1930
1578	YNSSDFTTNLNF-----VEPEGSDRSGTKLKOV-----	1605
1931	YNTKKSIKDIYKSLADEVVSITYKEMYKEMRDIELEAHQKATQDLIDEDKTDDAEKQKE	1990
1606	-----IQK-----QVNNNYADMGAYLTFWYDKNIIINQPNVITANI-----	1649
1991	LKERQDSIQKLTQDQINQYSLD-----DSEFGKSKVKELTEQLQEQBLDLDPLDK	2040
1650	-----VKELEDNTKLTAPITQWPNISGSKEKF-----YKPTVFGPNW	1688
2041	RESWKREALQDQLEKDEESINNKYDNLVNDERAFKLEBKIMNGKTIITAKQLNEFSKF	2100
1689	ENEN-SSMNSOQPTTWKIREGFPALQALKSSFDQKTRTFVLTITNAPLPLWKYCPGLGFQ	1747
2101	INTNWESIGKSSINNLIDKLKE---ASNALNTAVKGNITGKKVSSFAS---GGYTGTGLGA	2155
1748	GPNFKTQDWRILVFQND-----NOIAALRVQEQDRPEKSSDDKDKQK-----	1799
2156	GKLAFLDHDKELILNKTKTITANILDTVKAVRETAVD-----DSPKMGQGVKLADLIKK	2206
1800	EMFNSGNIRFVGVMQIQGPNTLWLPVNSSVIYDFYRGTDGSDNDVANLVAPQVKTITAF	1859
2207	-----GITSIPS-----LVPNVNQSL-----TNSLIPLNKKIEIPSKTIIAS	2243
1860	TNNAFNVPFKEFNISKKI	1877
2244	SGDKTILNLTTFHIDKLI	2261

```

1 Sequence 6, Application US/08480604A
2 Patent No. 5736139
3 GENERAL INFORMATION:
4 APPLICANT: KINK, JOHN A.
5 APPLICANT: THALLEY, BRUCE S.
6 APPLICANT: PADHYE, NISHA V.
7 APPLICANT: FIRCA, JOSEPH R.
8 APPLICANT: STAFFORD, DOUGLAS C.
9 TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
10 TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
11 NUMBER OF SEQUENCES: 32
12 CORRESPONDENCE ADDRESS:
13 ADDRESSEE: MEDLEN & CARROLL, LLP
14 STREET: 220 MONTGOMERY STREET, SUITE 2200
15 CITY: SAN FRANCISCO
16 STATE: CALIFORNIA
17 COUNTRY: UNITED STATES OF AMERICA
18 ZIP: 94104
19 COMPUTER READABLE FORM:
20 MEDIUM TYPE: Floppy disk
21 OPERATING SYSTEM: PC-DOS/MS-DOS
22 SOFTWARE: PatentIn Release #1.0, Version #1.30
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/08/480,604A
25 FILING DATE: 07-JUN-1995
26 CLASSIFICATION: 424
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: US 08/422,711
29 FILING DATE: 14-APR-1995
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: US 08/405,496
32 FILING DATE: 16-MAR-1995
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: US 08/329,154
35 FILING DATE: 25-OCT-1994
36 PRIOR APPLICATION DATA:
37 APPLICATION NUMBER: US 08/161,907
38 FILING DATE: 02-DEC-1993
39 PRIOR APPLICATION DATA:
40 APPLICATION NUMBER: US 07/985,321
41 FILING DATE: 04-DEC-1992
42 PRIOR APPLICATION DATA:
43 APPLICATION NUMBER: US 07/429,791
44 FILING DATE: 31-OCT-1989
45 ATTORNEY/AGENT INFORMATION:
46 NAME: INGOLIA, DIANE E.
47 REGISTRATION NUMBER: 40,027
48 REFERENCE/DOCKET NUMBER: OPDH-01763
49 TELECOMMUNICATION INFORMATION:
50 TELEPHONE: (415) 705-8410
51 TELEFAX: (415) 397-8338
52 INFORMATION FOR SEQ ID NO: 6:
53 SEQUENCE CHARACTERISTICS:
54 LENGTH: 2710 amino acids
55 TYPE: amino acid
56 TOPOLOGY: linear
57 MOLECULE TYPE: protein
58 US-08-480-604A-6

```

Db 185 FINTYKSOINKPTVPTIDDDIIKSHLVSEYARDETIVLESYRTNSLRKINSHGDIRAN-- 242
Qy 310 TSLFA-----NENDLKQDLDVLLDVSDFYGGQSETITSNSQVQVPASERS 356
Db 243 -SLFTEQELLNIYSQELLNRGNLAASDI VRLALKNFSG----- 281
Qy 357 LKORVKPKDOOKPRIEKPSLY-BYDAL-SFYSQJQELVSKPNSI----- 399
Db 282 -----VYLDVMDLPGIHSDFPKTISRPSISIGLDRWEMIKLEAIMK 321
Qy 400 -KOLVNATLARNRFLSLGKYNFLDDEDLASHL-DYFLVLSRAKIKOSSITKKLFIEPIKI 457
Db 322 YKKTINNTSEN-----FDKLDQQLKDNFKLIIESKSEKSI FSKL----- 362
Qy 458 SLKSSILGDQBPNTILFEKEVTFKLDNFRDVEIEKAPGLYPGVNBELOARQAQAS 517
Db 363 -----ENL-----NVSDEIKIAFAL-----GSVINGALISQGSY 393
Qy 518 EKEKSKGLKEFSQOKBENKAINNOE---GLEEDDNITE-----RLPENSPI 562
Db 394 -----LTLVAVIEQVKRYQFLNQHLPAPAESDNNFTDTTKIFHDSLFNSATAENSMF 445
Qy 563 -----QYQOENAGLGASDKPYMIK---DVONQRYV---LAKSOIQELIKADYTKLA 609
Db 446 LTKIAPYLQ---VGFMEARSTISLSPGAYASAYDFINLQENTIEKTLKASD----- 496
Qy 610 KLLSNRHTYINSLRLKEQLFDPNPRIPSSRDIEKAFVLDKTEKNKYWQIYSSASPVQON 669
Db 497 -LIEFPKPENNLSQLTEQ--EINSL--WSFDOASAKYQFEKYVRDYTGGSLSBNGVDNF 551
Qy 670 KWSLFGYRYLLG-----LDPKQTIHELKVLGQAGLOFEGYENLPSDFNLEDLK 719
Db 552 KNTALD-KNYLLNKKIPSNVVEEAGSKNYVHYIQL-QGDDISYEATCNL---FSKNPKN 606
Qy 720 NIRIKTLPFSQDNFKL---SLLDENNYDGEIKAPE-----FGLPLFL 760
Db 607 SIIOQRNNEKAKSVFLSDGESILELKY-----RIPERLKNKEKVKVTFIGHGKBEFN 661
Qy 761 PKELRNNSNGGGSQNSN-----SPWQEI-----ISOFKOQNSLNQOLQAFSTKI 807
Db 662 TSEPARLSVDSLSNEISSFLDTIKLIDISPKNVEVNLGNCNMFSDNFVEETYPGKLLISI 721
Qy 808 WEKIIGDENEFQNN-----RLQYKLLKDL---QESWINK-----TRNLY 845
Db 722 MDKITSTLFDVKNKSITIGANQYEVRIINSEGRKELLASHGKWINKEBEAIMSDLSKEYIF 781
Qy 846 WTYLGDGLKVKPKNNLEAKFRQISNLQELLTAFYTSAAIS-----NN----- 887
Db 782 FDSIDNKLKAKSKN---IPGLASISDGIKTLILDASVSPDTKFLNNLKLNTIESSIGD 836
Qy 888 WNYQDSGAKSTIIFEEIAEL---DPKVKKEGVADYVOLK-----FHYAIFDD---NAG 936
Db 837 YIYIEKLEPVKNI IHNISIDDLIDEFNLENVSDLEYELKLNLDKYLISFEDISKNNS 896
Qy 937 KPNQEVIRSSRTTYLTKSGK-----SKLEADPID 966
Db 897 TYSVRPINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGNLLDNIQLDHTS 956
Qy 967 QLNQAVKNAPLGQSPYLDTERFVQFKLATSLAVO----- 1002
Db 957 QVN--TLNAAFFIQSLIDYSSNKKVDNLNSTSVKQVLAQLFSTGLNTIYDSIQLVNLIS 1014
Qy 1003 -----HKQKEKT----- 1009
Db 1015 NAVNDTINVLPITTEGIPVSTILDGINLGAAIKELLDEHDPLLKKELEAKGVLAINMS 1074
Qy 1010 -----LPKLNNDGYTLIHKLKKPV---IPQISS 1037
Db 1075 LSGIAATVASIVGIGAEVITFLPIAGISAGISPLVNN---LTLHDKATSVVNYFNHLS 1132
Qy 1038 -----PEKDWEPGKLNQ-----GOSQNVNVTFGS- 1063

Db 1133 KKYGLPKTDDDKILVPIDDLVISEIDFNNNNSIKGLTCNILAMEGGSGHTVTGNIIDHFFSS 1192
Qy 1064 -IIESPYFTSNFQEDADLDQDGDSDRO-----GNSLNDQBEAGL 1102
Db 1193 PSISSHIPSLSIYSAIGIETENLDFSKIMMLPNAPSRVFWETGAVPGLRSLENDGTRL 1252
Qy 1103 LKOKLAILLGNQPIQYQQNDKEIE-----FEIINVEKVSSELSFR-----VEFK 1146
Db 1253 LDSIRDLYPGKFVYRFVAFPDYAITTLKPVYEDTNIKIKLDKDRNFIMPTITTTNEIRNK 1312
Qy 1147 LAKTLEDNGKTIKRVL-----SDETMSL-IWNTTIEKTPENSAY 1183
Db 1313 LSYSPFCAGGTYSLLSYSPISPTINILNSKDDLWIFNIDNEVRREISENGTIKKGLIKDV 1372
Qy 1184 PEVDFIKWVEQYDPRTPLAAKTFVLKFQDQIPVDGSGNISDK-----WL 1228
Db 1373 LSKIDIN-----KNKLI---GNQIDFSGDIDNKDORYIFLTCELDKISLI 1416
Qy 1229 ASIPLVTHOQML-----RLSPVVKVTIRELGLK-----TEQOQOQOQOQOQ 1269
Db 1417 IEINLVAKSYSLLSGDKNYLISNLSNTIEKINTLGLDSKNIAVNYTDESNNKYFGAISK 1476
Qy 1270 QPOKXAVRKEE---LETYNP-----KQEFNLN----- 1295
Db 1477 TSQKSIHYKKDSKNILEFYNDSTLEFNSKDFIAEDINVEFMKDDINTITGYKYVDNNTDK 1536
Qy 1296 -----PLTKAHLTLTSLNVNDPNYKIEDLVKNEAGDHLQAFSLRANNIKRLMNTPI 1349
Db 1537 SIDFSISLVSKNQVKNGLYLNESVYS-SYLDPFVKNSDGH-----NTSNFMNLFIDNI 1589
Qy 1350 TFADYNPFYNNEDWRSIDKYL-----NNKG-----NVSSHQOQAAA 1385
Db 1590 SF--WKLFGFENINF-VIDKYFTLVGKTNLGYVEFICDNNKNDIYFGEWKTSSSKSTIF 1646
Qy 1386 GGNQGS-----GLIORLANK-NIKPETFPTALIALKDRN 1417
Db 1647 SGNGRNVVVEPIYNPDTGEDISTSLDFSYPGLYIDRYINKVLIAPLYT-SLINI----- 1701
Qy 1418 NTNL--SNYSKIIMIKPK-----YLVERSIGVPWSTGLDGYIGSEQTKDGTSSSSQ 1467
Db 1702 NNTYSEYPEIILVNPNTFHKKNINLDBSSFEYKWSSTEGSDFILVRYLEESNKKILQ 1761
Qy 1468 Q---KGPDQDFIQALGKNTYHKGKGL-----SIRIFDPGNELAK----- 1505
Db 1762 KIRIKGI-----LSNTQSFNKSIDFKDIKKLSLGYIMSNFKSNSELDROHLG 1812
Qy 1506 IKDASNK---KGBEKLKSYDLFPQVLYNEVEKSPKIAKGTWNIHDOKEYPNPNQKLP 1561
Db 1813 FKIIDNKTYYYDEDSKLVKGLININNSLFPDPIEFNLVTGMQTI--NGKKY----- 1862
Qy 1562 ENYLNVLVNQVPKVTLY---NSSDFITN-----LPVBEPSGDRSGTKLKQVIOKQV 1610
Db 1863 --YFD--INTGAULTSVKIINGKHIFYNNDGVMOGLGVKPGDPGEYFAPANTQ----- 1911
Qy 1611 NNNYAD---WGSAYLT-----FWYDKN-----IITNQPNVITANIADVFIKDVKEL 1653
Db 1912 NNNIEQAIYVQSKFLTLNGKKYFDDNNSKAVTGWRIINNEKYFNPNNNA-IAAVGLQVI 1970
Qy 1654 EDNTKLIAPN---ITQWPNISGSKPEYKPTVFFGN 1687
Db 1971 DNKKYFNPDTAIIISKGWQTVNGSRYFDTDTAIAFN 2007

RESULT 20

US-08-405-496A-6

; Sequence 6, Application US/08405496A

; Patent No. 5919665

; GENERAL INFORMATION:

; APPLICANT: WILLIAMS, JAMES A.

; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:


```
Qy 1229 ASIPLVTHQML-----RLSPVVKTIRELGLK-----TEQOQQOQQOQQOQQ 1269
Db 1417 IEINLVAKSYLLSGDKNYLISNLSTIEKINTGLDLSKNIAYNYTDESNNKYFGAISK 1476
Qy 1270 QPOKKAVERKEE---LETNP-----KDEFNLTN----- 1295
Db 1477 TSQSIHYKDKSKNILEFYNDSTLEFNSKDFIAEDINVPKODINTITGKYVYDNNTK 1536
Qy 1296 -----PLTKAHLJTLNLVNDPNYKIEDUKVTKNAGDHQALAFSIRANNIKELMTPI 1349
Db 1537 SIDFSISLVSKNQVKNGLYLNESVYS--SYLDFVKNSDGH-----NTSNFNLFLDNI 1589
Qy 1350 TFADYNPFYNYEDWRSIDKYL-----NNKG-----NVSSHQOQAA 1385
Db 1590 SF--WKUFGFENINF-VIDKTYFTLVGKTNLGYVEFICNNKNIDYIGEWKTSKSTIP 1646
Qy 1386 GGNQGS-----GLIQRLNK-NIKPETFTPALIALKDRN 1417
Db 1647 SGNRNVVVEPIYNPDGTGEDISTSLDFSYEPLYGIDRVINKVLIAPDLYT-SLNI--- 1701
Qy 1418 NTNL--SNYSKLIIMIKPK-----YLVERSIGVPMWSTGLDGYIGSEQTKDGTSSSQ 1467
Db 1702 NTVYNSYEPYELIIVLNPTEHKKVINLNDSSPEYKMWSTEGSDFILVRYLEESKKILQ 1761
Qy 1468 Q---KGFPDQFIQALGLNTEYHGKGL-----SIRIFDPGNEELAK--- 1505
Db 1762 KIRIKGI-----LSNTQSFNMSIDFKIKKLSLGYIMSNFKSFNSENEIDRHLG 1812
Qy 1506 IKDASNK---KGBEKLKSYDLFQVNLNEYKKSPIAKGWTNIHPDQKEYPNPQKLP 1561
Db 1813 FKIDNKTYDYDESKLVKGLININNSLFYDFDIEFNLTGWTI--NGKXY----- 1862
Qy 1562 ENYLNVLNPKWVTLY--NSSDPITN-----LFVEPEGSDRGSGTKLKQVIOKQV 1610
Db 1863 --YFD--INTGAALTSYKIINGKHFFYFNNDGVMQLGVKFGPDGFEYFAPANTQ----- 1911
Qy 1611 NNNYAD---WGSALVT---FWVDKN-----IITNQPNVITANIADVFIKVKVEL 1653
Db 1912 NNNTEGQAIYQSKFELTNGKKYFYDNNKSVAWTGWRIIINNKYFNPNNNA-IAAVGLQVI 1970
Qy 1654 EDNTKLIAPN---ITQWPNISGSKEFYKPTVFFGN 1687
Db 1971 DNKKYFNPDTAISKGQWTVNGSRYYEDTDTAIAFN 2007
```

RESULT 21

US-08-915-136-6

Sequence 6, Application US/08915136

Patent No. 6290960

GENERAL INFORMATION:

APPLICANT: KINK, JOHN A.

APPLICANT: THALLEY, BRUCE S.

APPLICANT: PADHYE, NISHA V.

APPLICANT: FIRCA, JOSEPH R.

APPLICANT: STAFFORD, DOUGLAS C.

TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND

PREVENTION OF C. DIFFICILE DISEASE

TITLE OF INVENTION: 32

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/915,136

FILING DATE:

```
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-136-6
```

Query Match 2.6%; Score 250.5; DB 3; Length 2710;

Best Local Similarity 18.1%; Pred. No. 8.9e-08;

Matches 387; Conservative 274; Mismatches 647; Indels 829; Gaps 101;

```
Qy 180 LSNFSAQSIKPLATRSSDF--TAKLNQFNQDDELWV-----YLEKFFDLEAKA 227
Db 71 LSNLKKDILKEVILIKNSNTPVEK-----NLHFVWIGGEVSDIALEYIKOWADINA-EY 124
Qy 228 NIRLOTADFSPKGNLVDPPVYSP-----IRNPQ--NOKEWASDLN---QDQKT 271
Db 125 NIKLWYDSEAFVNTLKAIVESSTTEALQLLEEIQNPQPDNMKFYKRMFEIVYDQKR 184
Qy 272 VRLYLRTFS---POAKTILKQY---KY-KDETFLSS-----IDLKASNG 309
Db 185 FINYKSGINKPTVPTTIDDIKSHLVSEYNRDETVELSEYRTNSLRKINSNHHGIDIRAN-- 242
Qy 310 TSLFA-----NENDLKDQDLDVLLDVSDYFGQSGSETITSNSQVKVPASERS 356
Db 243 -SLTFEQLLNLYSQELLNRCNLAAASDI-VLLALKNFGG----- 281
Qy 357 LKDRVFKKQDQKPRIKFSLY-EYDAL-SFYSQLQELVSKPNSI- 399
Db 282 -----VYLDVDMLPGHISDLFTKISRPSIGLDRWEMIKLEATMK 321
Qy 400 -KDLVNATLARNLRFSLGKNFLPDDLASHU--DYFVLVSKAKIKQSSITKLFIELPIKI 457
Db 322 YKVINNYTSEN-----FDKLDQQLKDNFKLIIESKSEKSEIFSKL----- 362
Qy 458 SLKSSILGDQEPNLIKLFKEKVTFKLDNFRDVEKAFGLLYPGWNELEQAKQAQASAF 517
Db 363 -----ENL-----NVSDLEIKIAPAL-----GSVINQALTSKQSY 393
Qy 518 EKEKSKGLKEFSQKSENSKAINNQE---GLEBDDNITE-----RLPENSPI 562
Db 394 -----LTNLVIEQVKRYQFLNQHNLNPAIESDNNFTDTTKIFHDSLFNSATAENSMT 445
Qy 563 -----QYQENAGLQASPKPYMIK-----DVONQRY-----LAKSQIELIKAKDYTKLA 609
Db 446 LTKIAPYLQ-----VGFMEPEARSTISLGGPAGAYAYDFINLQENTIEKTLKASD----- 496
```

QY 610 KLSNRHTYNSIRLKEQLPDVNPRISSRDIEKAPVLDKTEKKNYQWISASPVFQ 669
DB 497 -LIEFAPPENLSQLTEQ--EINSL--WSPDQASAKYQKEKYVRDYTGGSLSNDGVDEN 551
QY 670 KWSLFGYRVLG-----LDPKQTHLVLKQKAGLOFEGYENLPSPDNEDLK 719
DB 552 KNTALD-KNYLLNNKIPSNVVEBAGSNVYHYIQL-QGDDISYEATCNL---FSKNPKN 606
QY 720 NIKITPLFSQKONFKL-----SLDDFNYYDGEIKAPE-----FGLPLEL 760
DB 607 SIIQRMNESAKSYFLSDGESILENKY-----RIPERLKNKEKVYTFIGHGDEEN 661
QY 761 PKELRNSNGSGQSN-----SPWEQBI-----ISQFKQNLNQDQLAQFSTKI 807
DB 662 TSEFARLSVDSLNEISSFLDTIKLDSIPKQVENVLLGCMFSDYFNVEETPGKLLSI 721
QY 808 WEKIIGDENEFDONN-----RLOYKLLDL--QESMINK-----TRDNLY 845
DB 722 MDKITSTLPVKNKSNITIGANQYEVINSEGRKELLAHSGKWINKEAMSDLSKEYIF 781
QY 846 WYLGDKLVKPKNNLEAKPROISNQLLTAFYTSAAALS-----NN-----887
DB 782 FDSIDNKLKAKSN-----IPGLASISEDIKTLLDASVPTKPFILNNLKNIIESSIGD 836
QY 888 WNYQSGAKSTIFBEIABL--DPKVKQKGVADVYOLK-----PHYAIGFDD--NAG 936
DB 837 YIYKLEPKNIIHNSIDDLDEFNLENVSDLYELKLANLDEKYLISFEDISKNS 896
QY 937 KFNQEVIRSSSTIYKTSK-----SKLEADTID 966
DB 897 TYSVRFINSGSVVETEKEIFSKYSEHITKEISTIKNSIITDVGNNLNDIQDHTS 956
QY 967 QLNQAVKNAPLQSGFVLDTERGVFOKLATS LAVO-----1002
DB 957 QVN--TLNAAFFQSLIDISSNKDVLNDLSTSVKQVLAQLFSTGLTYDSIQLVNLIS 1014
QY 1003 -----HKOKEKT-----1009
DB 1015 NAVNDTINVLPITTEGIPIVSTILDGINLGAAIKELLDEHDPLLKEAKVGVLAINMS 1074
QY 1010 -----LPKLANDGYTLIHDKLPV--IPOISS 1037
DB 1075 LSTAATVASIVGAEVTFILPIAGISAGIPSLVNE--LILHDKATSVVNVFNHLS 1132
QY 1038 -----PEKOWFEGLMKN-----GQSQNVNVSTFGS- 1063
DB 1133 KKYGLKTEDDKTLVPIDDLVISEIDFNNSIKLGTCNILAMEGGSGHTVTGNIDHFFS 1192
QY 1064 -ITESPYFTNFQEDADLDQDQDSDRQ-----GNNSLDNQEAGL 1102
DB 1193 PSISSHIPSLSIYSAIGIETENLDFSKIMMLNAPSFRVFWETGAVPGLRSLNDGTRL 1252
QY 1103 LKQKLAILLGNQFIQYQNDKEIE-----FEIINVEKVSLSFR-----VEFK 1146
DB 1253 LQSIIRLYPGKFWRFYAFDYAITTLKPVYEDTNIKIKLDKOTRNFIMPTITTEIRNK 1312
QY 1147 LAKTLEDNGKTIRVL-----SDETMSL-IVNTTIEKTPEMSAV 1183
DB 1313 LSVSFDGAGGTYSLLSSYPISNTINLSKDDLWIFNIDNEVREISIENTGIKKGKLIKDV 1372
QY 1184 PEVFTKWVEQYDPTPLAAKTFVLKFKDQIPVDGSGNISDK-----WL 1228
DB 1373 LSKIDIN-----KNKLIIT--GNQITDFSGDIDNKDRYIFLTCELDDKISLI 1416
QY 1229 ASIPLVHOOML-----RLSPVVKTTRELGLK-----TEQQOQQOQQOQQ 1269
DB 1417 IEINLVAKSYLLSGDKNVLNSLNTIEKINTLGLDSKNIAYNVTESSNNKYFGAISK 1476
QY 1270 QPOKKAVRKEE-----LETYNP-----KDEFNILN-----1295
DB 1477 TSQKSIHYKDSKNILEFYNDSTLFPNSKDFTAEDINVMKDDINTITGKYVVDNNTDK 1536

QY 1296 -----PLTKAURLTSLNVDNPNYKIEDLKVIKNEAGDHQALFSLRANNIKRLMNTPI 1349
DB 1537 SIDFSISLQKQVKNGLYLNESVYS--SYLDFVKNSDGH-----NTSNFMNLFLDNI 1589
QY 1350 TFADYNPPFYNNEDWRSIDKYL-----NNKG-----NVSSHQOQAA 1385
DB 1590 SP--WKLFGFENINP-VIDKYFTLVGKTNLGVVEFICDNNKIDIDYFGWKVSSSKSTIF 1646
QY 1386 GNGQGS-----GLIQLRNLK-NIKPETETFPALIALKORN 1417
DB 1647 SGNGRNVVVEPIYNPDTGEDISTSLDYSVEPLYGIDRYINKVLIAPDLVT-SLINI--- 1701
QY 1418 NTNUL--SNYSDKIIMKPK-----YLVERSIGVPSWSTGLDVGYSQTKDGTSSSQ 1467
DB 1702 NTNYSNEYEPIIIVLNPNTFHKVININLDSSSFYKWNSTEGSDFLVLYLEESNKILQ 1761
QY 1468 Q--KGFDQDFIQLGLKNTVEHGLGL-----SIRIFDGNELAK--- 1505
DB 1762 KIRIKI-----LSNTQSFNMSIDFKDKKLSLGYIMSNFKNSENELDRHLG 1812
QY 1506 IKDASNK-----KGEKLLKSYDLFKNYLNEYEKSPKIAKGTWNIHPDQKQYFNPKNQKL 1561
DB 1813 FKIIDNKTYVYDEDSKLVKGLININSLFYDPPIEBENLVGWTI--NGKYY----- 1862
QY 1562 ENYLNVLNQPWKVTLY--NSSDFITN-----LFVEPEGSDRGSTKUKQVIOKQV 1610
DB 1863 --YFD--INTGAALTSYKIIINGKHFYFNNDGVMQLGVFKGPDGFYFAPANTQ----- 1911
QY 1611 NNNYAD--WGSAYLT-----FWYDKN-----IITQNPVITANTADVFIDKVKEL 1653
DB 1912 NNNISQALVYOSKFTLNGKKYYFDNNKSAVTGRIINNEKYFNPNNNA-IAAVGLQVI 1970
QY 1654 EDNTKLIAPN---ITQWMPNISGSKEKYKPTVFFGN 1687
DB 1971 DNKKYFNPDTAISKGWQTVNGSRVYFDTUAIAPN 2007

RESULT 22

US-08-957-310-6

; Sequence 6, Application US/08957310

; Patent No. 6365158

; GENERAL INFORMATION:

; APPLICANT: Williams, James A.

; APPLICANT: Kink, John A.

; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES

; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE

; TITLE OF INVENTION: DISEASE

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Medien & Carroll

; STREET: 220 Montgomery Street, Suite 2200

; CITY: San Francisco

; STATE: California

; COUNTRY: United States of America

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/957,310

; FILING DATE: 23-OCT-1997

; CLASSIFICATION: 424

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/329,154

; FILING DATE: 24-OCT-1994

; APPLICATION NUMBER: US 08/161,907

; FILING DATE: 02-DEC-1993

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 07/985,321

; FILING DATE: 04-DEC-1992

; PRIORITY APPLICATION DATA:

```

; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-957-310-6

Query Match      2.6%; Score 250.5; DB 3; Length 2710;
Best Local Similarity 18.1%; Pred. No. 8.9e-08;
Matches 387; Conservative 274; Mismatches 647; Indels 829; Gaps 101;

Qy 180 LSNFSAOSIKPLALTRSDSF--IAKLQNFNNQDLWV-----YLEKFPDLALKA 227
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 LSNLKKOILKEVILKNSNTSPVEK-----NLHFVWIGGEVSDIALEVIKQWADINA-EY 124

Qy 228 NIRLOTADFSPEKNLVDVPFYVSF-----IRNPQ--NOKEWASDLN---QDQKT 271
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 NIKLWYSEAFVNTLKKAVI SESSSTTEALQLEEEIQNPQDNMFKYKGRMEFYDRQR 184

Qy 272 VRLVLRTEFS---POAKTILKDY---KY-KDETFLSS-----IDLKASNG 309
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 185 FINYKQINQPTVPTDIDIIKSHLVSEYNRDETVELESYRNSLRKNSNHGIDIRAN-- 242

Qy 310 TSLFA-----NENDLKQDLVDLLVDVFGGOSSETITSNSQKVPVPSERS 356
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 243 -SLFTEQELLNYSQELNLRGNLAASDI VRLALKNFG-----281

Qy 357 LKDRVKFKDQOKPRIEKFSLY-EVDAL-SFYSQLOELVSKPNSI-----399
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 282 -----VYLDVMDLPGIHSDLFKTI SRPSSIGLDRWEMIKLEAIMK 321

Qy 400 -KDLVNATLARNLRFSLGKYNFLPDLLASHL-DYVFLVSKAKIKOSSITKKLTLPKI 457
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 322 YKKYINNTSEN-----PKLQDQKDNFKLIIESEKSEIFSKL-----362

Qy 458 SLKSSILGQBPNIKTLPKEVTFKLDNFRDVEIEKAFGLLYPGVNEELEQAKARASF 517
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 363 -----ENL-----NVSDLEIKIAFAL-----GSVINQALISKQGSY 393

Qy 518 EKEKSKGLKEFSQOKEKNSKAINQOE---GLEEDDNITE-----RLPENSPI 562
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 394 -----LTNLVIEQVKNRYOFLNQHNLNPAIESDNFTDTTKI PHDSLFPNSATAENS MF 445

Qy 563 -----QYQENAGLGPDPKPMYK-----DVQNRYY---LAKSOIOLIKAKDYTKLA 609
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 446 LTKIAPYLQ-----VGFMEPEARSTISLSPGAVASAYDFINLQENTIEKTLKASD-----496

Qy 610 KLLSNRHTYINSLRLEQLDFVNPRI PPSRDIEKAKFLDKTEKNKYWQIVSSASPVFQN 669
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 497 -LIEKFPFNLSQLTEQ--EINSL--WSFDQASAKYQFVKYRVDTYTGSSLEDNGVDNF 551

Qy 670 KWSLFGYVRYLLG-----LDPKQTIHLVKGKAGLOFPGYENLPSDFNLEDLK 719
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 552 KNTALD-KNYLLNNKIPSNVBEAGSKNYVHYIQL-QGDDDISYBATCNL---FSKNPKN 606

Qy 720 NRIKTPFLSQDNFKL-----SLLDNNYDGETIKAPE-----FGLPLFL 760
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 607 SLIIQRMNESAKSYFLSDGGSILELNKY-----RIPERLKNKPKYKVTFFIGHKDEFN 661

Qy 761 PKELARNSSNGSQNSN-----SPWEQEI-----ISQPKQNLNQDLAQFSTKI 807
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 662 TSEFARLSVDSLNEISSFLDTIKLDISPKNVEVNLGCMNFSYDFNVEETYPGKLLLSI 721

808 WEKIIGDENEFDONN-----RLQYKLLKDL--QESWINK-----TRDNLV 845
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
722 MDKITSTLPDVNKNSTIGANQVEYINSEGRKELLASHGKWINKEEAIMS DLSKEYIF 781
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
846 WTYLGDKLVKPKQNLLEAKFRQISNLOELLTAPYTSAAALS-----NN-----887
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
782 FDSIDNKLKAKSKN-----IPGLASISEDIKTLLLDASVSPDTKFLNNLKLNISSIGD 836
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
888 WNYQDSGAKSTIIPBEIAEL--DPKVKEKGVADVQLK-----PHYAIGFDD--NAG 936
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
837 YIYIEKLEPVKNIIHNSIDDLIDEFNLEENVSDELIELKLNLDKYLISPEDISKNS 896
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
937 KPNQEVIRSSRTIYLKTSK-----SKLEADTID 966
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
897 TYSVRFINKSGESVVVETEKEIFSKVSEHITKEISTIKNSIITDVNGLLDNIQLDHTS 956
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
967 QLNQAVKNAPLGLOSFYLDTEREGVFQKLAISLAVQ-----1002
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
957 QVN--TINAAFFIQSLIDYSSNKDVLNLDLSTSVKQVLAQLFSTGLNTIYDSIQLVNLIS 1014
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1003 -----HKQKEKT-----1009

1015 NAVNDTINVLPITTEGIPVSTILDGINLGAALKELLDEHDP LKKELEAKVGVLAINMS 1074
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1010 -----LPKLANNDGYTLIHDKLPV--IPOISS 1037
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1075 LSTAATVASIVGIGAEVTIFELLPIAGISAGISPSLVNNE--LILHDKATSVVVPNHLSES 1132
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1038 -----PEKOWFEKLNQ-----QSQNVNVSTFGS- 1063
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1133 KKYGPLKTEDKILVPIDDILVISEIDFNNSNIKLGTCNILAMEGGSGHTVTGNIDHFSS 1192
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1064 -ITESPFSNFQEDADLDDGODDSRQ--GNNSLDNQBEAGL 1102
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1193 PSISHTIPSIUSIYSAIGIETENLDFSKIMLWNPASRVFWETGAVPGLRSLNDGTRL 1252
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1103 LKQKLAILLGNQFTQYYQNDKEIE-----FEIINVEKVSLSFR-----VEFK 1146
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1253 LDSIRDLYPGKFYWRFAFDYAITLLKPVVEDTNIKIKLDKOTNFIMPITTTNEIRNK 1312
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1147 LAKTLENGKTIKRVL-----SDETMSL-IYNTTIEKTPMSAV 1183
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1313 LSVSFDGAGGTYSLLSSYPISSTINLSKDBLWTFINIDNEVREISIENTGIKKGLKIDV 1372
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1184 PEVPTKWEQYDRTPLAAKTKFVLAKQIPVDGSGNISDK-----WL 1228
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1373 LSKIDIN-----KNKLI--GNQTI DFGSDINDKDRYIFLTCELDOKTSLI 1416
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1229 ASIPLVIHQWL-----RLSPVVKTIRELGLK-----TEQOQOQOQOQOQOQ 1269
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1417 IEINLVAKSYLLSGDKNYLISNLSNTIEKINTLGLDSKNIAANYTDESNNKYFGAISK 1476
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1270 QPKKAVRKEE---LETYNP-----KDEFNINL-----1295
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1477 TSQKSIHYKDKSKNILEFYNDSTLEFNSKDFIAEDINVFMKDDINTITGKYVDNNTDK 1536
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1296 -----PLTKAHLRTLNLNVNDPNYKIEDLKVINEAGDHOLAFSLRANNIKRLMNTPI 1349
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1537 SIDFSISLVSKQVKNGLYLNESVYS--SYLDFVKNSDGHH-----NTSFMNLFNDNI 1589
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1350 TFADYNPFYVNEWRSIDKYL-----NNKG-----NVSSHQOQAAA 1385
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1590 SF--WKLFGFENINF-VIDKYFTLVGKTNLGVVEFICDNNKNIDIVFGEWKTSKSTIF 1646
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1386 GGNQCS-----GLIQRLNK-NIKPEFTTALTALKDRN 1417
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1647 SGNGRNVVEPIYNPDTGEDISTSLDSFSEPLYGIDRYINKVLAPDLYT-SLINI-----1701
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1418 NTLN--SNYSDKIIMIKPK-----YLVERSIGVPWSTGLDYGISEQTKDGTSSSQ 1467
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1702 NNTYVSNEYPIEIVLNPNTTHKKVNIINLDSSSFYKWSSTEGSFILVRYLEESNKILQ 1761
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1468 Q---KGFQDQDFIQALGLKXNTEYHGKGLG-----SIRIFDPGNELAK-----1505
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 1762 KIRIKI-----LSNTQSFNKSIDFKDKIKLSGLGYMSNFKSFNSSELDHRDLG 1812
Qy 1506 IKDASNK-----KGEEKLLKSYDLFKYLYNEYEKKSPKIAKGWTHIHDPDOKEYPNPKLP 1561
Db 1813 FKIIDNKTYIYDEDSKLVKGLININSLFFYDFDPIEFNLVTGMQTI--NGKKY----- 1862
Qy 1562 ENYLNVLNQPWKVTLY---NSSDFTN-----LFVEPEGSDRGSGTKLKQVQKQV 1610
Db 1863 --YFD--INTGAALTYSKIINGKHFFVNDGVMLGVFKGPDGFYFAPANTQ----- 1911
Qy 1611 NNNYAD-----WGSAYLT-----FWYDKN-----IITNQPVITANTADVFKDKV 1653
Db 1912 NNNIEGQAIYQSKFLTLNKKYFYDNNNSKAVTGWRRIINNEKYFNPNA-IAAVGLQVI 1970
Qy 1654 EDNTKLIAPN---ITOWMPNISGKFKYKPTVFFGN 1687
Db 1971 DNKKYFNPDTAISKGWTVNGSRYVDFDTAIAFN 2007

RESULT 23
US-10-011-366-6
; Sequence 6, Application US/10011366
; Patent No. 6573003
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; INVENTOR: Kink, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/011,366
; FILING DATE: 16-No. 6573003-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-011-366-6

Query Match 2.6%; Score 250.5; DB 4; Length 2710;
Best Local Similarity 18.1%; Pred. No. 8.9e-08;
Matches 387; Conservative 274; Mismatches 647; Indels 829; Gaps 101;

Qy 180 LSNFSAQSIKPLATRSSDF--IAKLNQFNQDELWV-----YLEKPPDLAALKA 227
Db 71 LSNLKDILKEVILKNSNTSPVEK-----NLHFVWIGGEVDIALEYIKQWADINA-EY 124
Qy 228 NLRLOTADSPFKGNLVDPPVYSF-----IRNPQ--NQEWAADLN--QDOKT 271
Db 125 NIKLWDSAPLWTLKKAIVESSTEALQLEBEELQNFQDNMKFYKRMFEIYDQKR 184
Qy 272 VRLYRTEFS---POAKTILKDY---KY-KDETFLSS-----IDLKASNG 309
Db 185 FINYKSGINKPTVPTIDDIKSHLVSEYNRDETVLSEYRTNSLRKINSNHGDIRAN-- 242
Qy 310 TSLFA-----NENDLKQDLVDLDVSDYFCGQSETITSNQVPPVPSERS 356
Db 243 -SLFTEQELLNYSOELLNRGNLAASDVRLLAKNFGG----- 281
Qy 357 LKDRVFKKDKQOKPRIEKFSLY-EYDAL--SFYSQQLVSKPNSI----- 399
Db 282 -----VYLDVDMPLGIIHSDLFKTISRPFSSIGLDRWEMIKLEAIMK 321
Qy 400 -KDLVNATLARNLPSGLKYNFLFDDLAHL--DYPLVSKAKIKQSSITKFLFIELPIKI 457
Db 322 YKKYINNYTSEN-----FDKLDQOLKDNFKLIIESKSEKSEIFSKL----- 362
Qy 458 SLKSSILGQEQENIKTLPEKEVTFKLDNFRDVEIEKAFGLLYPGVNEELEQAKQARAP 517
Db 363 -----ENL-----NVSDLEIKIAFAL-----GSVINQALISQGSY 393
Qy 518 EKEKSKKGLKFEFSQOKEENSKAINNQE---GLEEDDNITE-----RLPENSPI 562
Db 394 -----LTNLVIEQVKRYQFLQHLNPAIESDNNFTDTTKIHFDSLFSNATENSMP 445
Qy 563 -----OYQENAGLGASPKPYMIK-----DVQNORYY-----LAKSOIELIKAKDYTKLA 609
Db 446 LTKIAPYLQ-----VGFMPEARSTISLSPGAVASAYYDFINLQENTIEKTLKASD----- 496
Qy 610 KLLSNRHTYNISLRKEQLFDVNPRIPSRDIKAKFVLDTKTEKNKYWQIYSSASPVFQN 669
Db 497 -LIEFKFPENNLSQLTEQ--EINSL--WSFQASAKYQPEKYVRDYVTGSGLSBEDNGVDPN 551
Qy 670 KMSLFQYRYLLG-----LDPKQTIHELKVLGKAGLQFEGYENLPSDFNLDELK 719
Db 552 KNTALD-KNYLLNNKIPSNVVEAGSKNVYHVIQL-QGDDISYEATCNL---FSKNPKN 606
Qy 720 NRIKTPLFQKDNFKL-----SLDPFNYYDGEIKAPE-----FGLPLFL 760
Db 607 STIIQRNMNESAKSYFLSDGGSILELKY-----RIPERLKNKEKVKVTFIGHGDEFN 661
Qy 761 PKELARNSSNGGSONS-----SPWQEI-----ISQFKDQNLNQDQLAQSFSTKI 807
Db 662 TSEFARLSVDSLNSNEISSFLDTIKDISPKVNEVNLGCMFSDYDNVEETPGKLLLSI 721
Qy 808 WEKIIIGDENEFQNN-----RLQYKLLKDL---QESWINK-----TRDNLY 845
Db 722 MDKITSTLPDVNKNSTITIGANQVEYRINSEGRKELLAHSGKWINKEAIAWSDLSKEYIF 781
Qy 846 WYLGDKLVKVPKNLEAKFROISNLQELLTFATYSAALS-----NN----- 887
Db 782 FDSIDNKLKAKSKN-----IPGLASISEDIKTLLDASVPDTKFLNNLKLNISSIGD 836
Qy 888 WNYQDSGAKSTIIFEEIAEL--DPKVKVKGVVYOLK-----PHYAIGFDD---NAG 936
Db 837 YIYEKLEPVKNIIHNSIDDLDEFNLENVSDLEYELKLNLDKEYLISFEDISKNS 896
Qy 937 KFNQEVY-RSSRTIYLTSGK-----SKLEADTID 966
Db 897 TYSVRFINKSGESVYVETEKEIFSKYSHITKSTIKNSITDVNGNLLDNLIQDHTS 956

Db 185 FINYKQINQKPTVPTIDDIKSHLVSEYNRDETIVLESYRTNSLRKINSNHGDIRAN-- 242
QY 310 TSLFA-----NENDLKQDQDVLDDVDFYFGQSETITNSQVKVPASERS 356
Db 243 -SLFTQELLNYSQELLNRGNLAASDIVRLALKNFG----- 281
QY 357 LKDRVFKDQKQPRIEKFSY--EYDAL--SFYSQQLVSKPNSI----- 399
Db 282 -----VYLDVMDLPGIHSDLFKTISRPSISGLDRWEMIKLEAIMK 321
QY 400 -KDLVNATLARNLRFSLGKNYFIPDDLAHL--DYFVLVSKAKIKQSIYTKLFIELPIKI 457
Db 322 YKYYINNYTSEN-----FDKLDQQLKDNFKLIIESKSEIIFSKL----- 362
QY 458 SLKSSILGDQEPNIKTLFEKEVTFKLDNFRDVEIERKAFGLLYPGVNEEQLQAKAQRASF 517
Db 363 -----ENL-----NVSDLEIKIAFAL-----GSVINQALISKQSY 393
QY 518 EKEKSKGLKEFQOQKEENSKAINQOE--GLEEDDNITE-----RLPNSPI 562
Db 394 -----LTLNVIEQVGRYQFLNHLNPAIESDNNFTDITKI FHDLSLFNSATAENSMF 445
QY 563 -----QYQOENAGLGASDPKPYMIK-----DVQNRYY-----LAKSOIQELIKAKDYTKLA 609
Db 446 LTKIAPYLO-----VGFMEPEARSTISLSPGAYASAYYDFINLQENTIEKTLKASD----- 496
QY 610 KLASNRHTYNSLRLEKQLPDVPNPRIPSSRDIEKAKFVLDTKTEKNYWOIYSSASVPFON 669
Db 497 -LIEFKFPENNLSQLTEQ--EINSL--WSPDQASAKYQEKYVRYDTGGSLSLEDNGVDFN 551
QY 670 KWSLFGYRYLLG-----LDPKQTHLHVKLQKAGLQFEGEYNLPSPDNLEDLK 719
Db 552 KNTALD--KNYLLNKNKIPSNVVEBAGSKYVHYIIQL--QGDDISYEATCNL--FSKNPKN 606
QY 720 NIKIKTFLFQKDNFKL-----SLDDPNYVDEIEKAPE-----FGLPLFL 760
Db 607 SIIIQRMNMSAKSYFLSDGSESILENKY-----RIPERLKNKEKVKVTFIIGHGDEFN 661
QY 761 PKELRRNSSGSGSONS-----SPWEQBI-----ISQFKDQNLNQDLAQFSTKI 807
Db 662 TSFARLSDVLSNEISFLDTIKDISPKVNEVNLGCMNFSYDFNVEETYPCKLLLSI 721
QY 808 WEKIIIGDENBFDQNN-----RLQYKLLKDL--QESWINK-----TRDNLY 845
Db 722 MDKITSTLPDVNKNSTIGANQVEYRINSEGRKELLASHGKWKINKEEAIMSDLSSKEYIF 781
QY 846 WTVLGDKLKVPKNLEAKERQISNLOELLTAFYTSAAALS-----NN----- 887
Db 782 FDSIDNKLKAKSKN-----IPGLASISEDIKTLILDASVSPDTKFTLNNLKLNISSIGD 836
QY 888 WNYQDSGAKSTIIFEBIAEL--DPKVEKVGADVYOLK-----PHYAIGFDD--NAG 936
Db 837 YIYKLEPKNVLIHNSIDDLIDEFNLLENVSDLELKLNNLDEKYLISFEDISKNS 896
QY 937 KPNQEVIRSSRTIYKLTSGK-----SKLEADTID 966
Db 897 TYSVRPFINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVGNLLDNIQDLHTS 956
QY 967 QLNQAVKNAPLGLQSFYLDTERFGVQKLATSVAQ----- 1002
Db 957 QVN--TLNAAFFIQSLIDYSSNKDVLNLDLSTSVKQVLAQLFSTGLNTIYDSIQLVNLIS 1014
QY 1003 -----HKOKEKT----- 1009
Db 1015 NAVNDTNVLPTEGIPVISTILDGINLGAAILKEILLDEHDPLLKLEAKVGVLAINMS 1074
QY 1010 -----LPKKLANNDGYTLIHDKLKXKPV--IPOISS 1037
Db 1075 LSLAATVASIVGIGAEVTIFLLPIAGISAGIPSLVNN--LILHDKRATSVVNYFNHLS 1132
QY 1038 -----PEKOWFEGLKON-----QSQNVNVSTFGS- 1063

Db 1133 KKGPKLTEDDKILVIPIDDLVISEIDFNNSIKLGTCLNLAEGSGCHTVTGNIDHFFPS 1192
QY 1064 -IIESPYSTNFQEDADLDQDQDSDRQ-----GNSLNDQEAGL 1102
Db 1193 PSISSHIPSLSYSAIGIETENLDFSCKIMLNPASRVFWMETGAVPGLRSLENDGTRL 1252
QY 1103 LKQKLAILLGNQFIQYQONDKIEB-----FPIINVEKVSLSFR-----VEPK 1146
Db 1253 LDSIRDLYPGKFWRYFAFFDYAITTLKPEVYEDTNIKIKLDKDRNFIMPTITTNIRNK 1312
QY 1147 LAKTLEDNGKTRVL-----SDETMSL--IVNTTIEKTPMSAV 1183
Db 1313 LYSYFDGAGTYSLLSSYPISNTINLSKDDLIWFINIDNEVREISIENTIKKGLIKDV 1372
QY 1184 PEVFDTKWVEYDPRTPLAAKTKFVLKFKDQIPVDSGSGNISDK-----WL 1228
Db 1373 LSKIDIN-----KNKLII--GNQTIIDFGSDIDNKDRYIFLTCBLDDDKISLI 1416
QY 1229 ASIPLVIHOOML-----RLSPVVKTRBELGLK-----TEOQOQOQOQOQOQ 1269
Db 1417 IBINLVAKSYSLLSGDKNYLISNLTIEKINTLGLDLSKNIAYNYTDENNKYFGAISK 1476
QY 1270 QPKKAVRKEEB--LETYNP-----KDEENILN----- 1295
Db 1477 TSQKSIHYKDSKNILEFYNDSTLEFNSKOFIAEDINVPMKDDINTIITGKYVVDNNTDK 1536
QY 1296 -----PLTKAHLRLTSLNVLNNDPNYKIBDLKVIKNEAGDHQLAFSLRANNIKRLMPTI 1349
Db 1537 SIDFSISLVSKNQKVGNGLYLNEYSYS--SYLDFVKNSDGHH-----NTSNFNLPLDNI 1589
QY 1350 TPADYNPPFPYNNEDWRSIDKYL-----NNKG-----NVSSHQOQAA 1385
Db 1590 SF--WKLFGFENINF--VIDKYFTLVGKTNLGVVEFICDNNKNIDYFGEWKTSSSKSTIF 1646
QY 1386 GGNQGS-----GLIQRANK--NIKPETFTPALIALKDRN 1417
Db 1647 SGNRNVVVEPIYNPDGTGEDISTSLDSFSEPLYGIDRYINKVLLIADPLYT--SLINI----- 1701
QY 1418 NTNLL--SNYSDKIIMIKPK-----YLVBSRISGVPMWSTGLDYGISGOTKDGTSSSSQ 1467
Db 1702 NTNYSNEYYPEIIVLPNPTFHKKVNIINLDSSEFYKWKSTEGSDFLVRYLESNKILQ 1761
QY 1468 Q--KGFDQDFQALGLKNTYHGKGL-----STRIFDPGHLELAK----- 1505
Db 1762 KIRIKGI-----LSNTQSFNKSIDFDIKKLSLGYIMNSFKSFENSELDRHLG 1812
QY 1506 IKDANK-----KGEKLLKSYDLFKNYLNEYKKGPKIAKGWNTIHPDOKEYPNPQKLP 1561
Db 1813 FKIDNKTYVDEDESKLVKGLININNSLFYDPDIEFNLVTGHQTI--NGKYY----- 1862
QY 1562 ENYLNVLNQPWKVTLY--NSSDFITN-----LFVEPEGSDRGSGTKLKQVIQKV 1610
Db 1863 --YFD--INTGAALTSYKIINGKHFFYFNDQVNLGVKFGPDGFYFAPANTQ----- 1911
QY 1611 NNNYAD--WGSAYLT-----FWYDKN-----IINQPNVITANTADYFIKDVKEL 1653
Db 1912 NNNIEQAIYQOSKFTLTLNGKYYFDNNSKAVTWGRINIINNEKYFNPNNNA--IAAVGLQVI 1970
QY 1654 EDNTKLIAPN---ITQWPNISGSKKFKYKPTVFFGN 1687
Db 1971 DNNKYFNPDTAILSKGWOTVNGSYFYFDTDTAIAFN 2007

RESULT 25

US-08-603-753D-4

; Sequence 4, Application US/08603753D

; Patent No. 5851857

; GENERAL INFORMATION:

; APPLICANT: HOLT, JEFFREY T.

; APPLICANT: JENSEN, ROY A.

; APPLICANT: PAGE, DAVID L.

; APPLICANT: KING, MARY-CLAIRE

; APPLICANT: SZABO, CILLIA I.

APPLICANT: JETTON, THOMAS L.
APPLICANT: ROBINSON-BENION, CHERYL L.
APPLICANT: THOMPSON, MARILYN E.
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARLES A. TAYLOR, JR.
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
CITY: DURHAM
STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27707

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
COMPUTER: IBM PC/XT/AT compatible
OPERATING SYSTEM: Windows 3.1
SOFTWARE: WORD PERFECT 6.1 and ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/603,753D
FILING DATE: 20 FEB 1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/373,799
FILING DATE: 17 JAN 1995
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/2
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens sapiens
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: female breast
CELL TYPE: normal breast tissue
CELL LINE: HMEC
ORGANELLE: no

FEATURE:
NAME/KEY: BRCA2 protein
LOCATION: 1 to 3418; Genbank locus HSU43746
IDENTIFICATION METHOD:
OTHER INFORMATION: BRCA2 protein has a negative
OTHER INFORMATION: regulatory effect on growth of human mammary cells.
PUBLICATION INFORMATION:
AUTHORS: Wooster, R. et al.
TITLE: Identification of the breast cancer
TITLE: susceptibility gene BRCA2
JOURNAL: Nature
VOLUME: 379
PAGES: 789-792
DATE: 1995
RELEVANT RESIDUES IN SEQ ID NO: 4: granin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344

US-08-603-753D-4
Query Match 2.6%; Score 250.5; DB 2; Length 3418;
Best Local Similarity 18.6%; Pred. No. 1.3e-07;
Matches 387; Conservative 280; Mismatches 716; Indels 693; Gaps 96;

QY	85	EFTDVSFKDFLTNNGRTVLEIPKQYQVWIEBSPB-----DDKERF	126
DB	1126	EFTQF-RKPSYILQ--KSTFEVENQMTILKTTSECRDADLHVIMNAPSIGQVDSKQF	1182
QY	127	RLGFHLKEKLEDGNIQAQSATKFIYLLPLDMPKAALGOYSYIVDNK-----	171
DB	1183	EGTVEIKR-----KFAGLLKNDCKNSAGS---YLTDENEVGFRGYSAHGTLK	1227
QY	172	-----FNNLIHPLSNFSAQSIKPIALTRSS--DFIAKLQNFQNODELWVYL	216
DB	1228	NVSTEALQKAVKLFSD--IENISEETSAEVHPISLSSCKCHDSVWSFMKIENHNDKTVE	1285
QY	217	EKFFDLEALKANIRLOTADSPFEKGNLVDPVVSFIRNPONQ-----	258
DB	1286	KNNKCQILQNNIEMT-----GTFVEITEYKRNTEENEDNKYTAASNSHNLEFDG	1338
QY	259	-----KEMASDLNQDQKTVRLYLRTFPSPQAKTILKDYKYKDTFTFSSIDLK-	305
DB	1339	SDSSKNDTVCIHKDETDLLFTDQHNICLSLGGQFMKEGNTQIKB-DLSDLTFLLEVAKAQE	1397
QY	306	-----ASNGTSLFA--NENDLKOQLOVDLDVDFGQSGSETITSNSQKVPVPSERSLK	358
DB	1398	ACHGNTSNKEQLTATKTEQNIKD-----FETSD-----TFQTAGSKNISVAKSELEN	1444
QY	359	DRVKFKKQKQKPRIEKFSL-----YEYDALSPY-----SQL	389
DB	1445	KVNF-PDQRPBELHNSLSLSELHSDIRKKNMILSYEETDIVKHKLKESVPVGTGNQL	1503
QY	390	QELVSKPENSIKDLVNATL-----ARNLRFSLGKYNPLF-----	423
DB	1504	VTFOGQPERDEKIKEPTLLGFHTAGSKKVIKAKESLDKVKNLDEKQGTSEITFSFHOW	1563
QY	424	-----DLASHLDYFVLSKAKIKQSSITKK---LFLELPKISLSKSSILGDQ	467
DB	1564	AKTKYREACKDLELACETIETAAPKCKEQMNSLNDKKNLVSLETVPVKLLSDNLCRQ	1623
QY	468	EPNITLFEKEVTFKLDNPRDVEIEKAFGLLYPGWNEE-----LQAKAQRASPEKBS-	522
DB	1624	TENLKT--SKSIFLKVKVHENVKEKETAKSPATCTYNQSPYSVIENSALAFYTSCSRKTSV	1681
QY	523	-----KRGKLE--FSQOKEENSKAINNOGLEDDNITERLPENSPITQYOQENAGLG	572
DB	1682	SQTSLLKAKKWLREGIFDGOPIRINTA-----DYGVNLYENNSNSTIAEN-----	1727
QY	573	ASPDPMIKDVQVORYLAKSQIOELIKADYTKLAKLLSNRHTYINISLKLQLPDVN	632
DB	1728	---DKNHL---SEKQDTYLSNS-----MSNSYSYH-----SDEVYNS	1760
QY	633	PRIPSSRDIEKAKFVKDKTEKNKYQIYSSASPVFNKWSLFGYRYVLLGLDPRQTIHEL	692
DB	1761	GYLSKNKLDGIEPLVKNVEDQK-----NTSFKVISNVKDNAY-----PQTVNE-	1806
QY	693	VKLQKAGLOPGEYENLPSDFNLBNLKNIRIKTLPFSQKNFKLSLLDFNNYDGEIKAP	752
DB	1807	-----DICVEEL--VTSSPCKNKNAIKLSISNSNNF---EVGPP	1842
QY	753	BFGPLPLFLPKELRRNSNSGSGNSNSPWEDEIISQFKD-----QNLNQDOLA	801
DB	1843	AFRI-----ASGKIRLCS---HETIKKVKDIFDTSFKVIKENNENKSKIC	1885
QY	802	QFSTKIM-----EKIIGDENEFQDNRLQVKKLLKQLQESWINKTRDNLYWTVLGD	851
DB	1886	Q--TKIMAGCYEALDDSEDLHNSLDNDECSMHSHKVPADIQSEILQHNQM--SGLEK	1941
QY	852	KLKVKPKN-NLEAKPRQISNLQELLTAPYTSAAALSNMNNYYQDSGAKST-----	899
DB	1942	VSKISPCDVSLTSDICKCSIGKL---HKSVSANTCGIFSTASGKSVQVSDASLQVAR	1997
QY	900	IIFBEIAELDPKVKVKV-----GADYVQLKFHVAIGFDD---NAGFNQVQVSSSRTI	950
DB	1998	QVFSEIEDSTKQVFSKVLFKSNEHSDQLTRENTAIRPTEHLISQKGFPSVNVNSSFASFQ	2057

[illegible]

RESULT 30

```

US-09-150-867-1
; Sequence 1, Application US/09150867
; Patent No. 6645748
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth W.
; APPLICANT: Sakowicz, Roman
; APPLICANT: Goldstein, Lawrence S.B.
; APPLICANT: Cleveland, Don W.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for
; TITLE OF INVENTION: Chromosome Congression
; FILE REFERENCE: 18557C-000110US
; CURRENT APPLICATION NUMBER: US/09/150,867
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/058,645
; EARLIER FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2954
; TYPE: PRT
; ORGANISM: Xenopus sp.
; FEATURE:
; OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
; OTHER INFORMATION: member of the kinesin superfamily of microtubule
; OTHER INFORMATION: motor proteins
; FEATURE:

```

```
QY 713 FNL-----EDKNIRIKTPLFSQKDNFKLSLLDNFNNYDGEIKAPERGLPLFLPKELRRNS 768
Db 1233 EKLTQLOEEMKNITI-----ERNELQTNFEDLKAEBDS-----LKQDLSENI 1274
QY 769 SNGSGQNSNSPWEQ-----IISOPKQONLS-----NODLAQFSTKTWEKI-IG 813
Db 1275 EQSIETQDELRQAQEBUREQOKLVDSFRQULLDCSVOISSPNHDAVAN-----QEKVSLG 1329
QY 814 DENEFDQNNRLQYKLLKDLQESWINTKRDNLW-----TYLGDKLKVK 856
Db 1330 EVN-----SLQSEMLRGERDELQTSCKALVSELLRAHVKSVEGENLEIT 1375
QY 857 PK-NLLEAKFRQISNIOELLTATYTSAAUSNNNNYQDSQAKSTIIFEEIAELDPKVK 915
Db 1376 KKLNGLEKEITLGSSESVLKS-----MLENLKEDNNKLKEQ 1412
QY 916 VGADVQLKPHYAIGPDNDNAGKFNQEVIRSSRTIYKTSKSKLEADTIDQLNQAVKA 975
Db 1413 --AEEYSSK-----ENQFSLEEV-----FSGSOKL-VDEIEVLKAQKAA 1449
QY 976 PLGLQSFYLDTERFGVQ-----KLATSLAVQHKQKE-----KTLPKKLAN 1016
Db 1450 EERLE--IKDRDYFELVQTANTNLVEGKLETPLOADHEEDSIDRSEMEIKVLGEKLER 1507
QY 1017 DGYTLIHDKUKKVPVIOISSPE-----KDWFEGLNQNGQONV-----1056
Db 1508 NOYLL--ERLQEBKL-ELSNKLEILOKEMETSVLLKDDLOQKL-ESLSNENIILKENIDT 1563
QY 1057 -----NVSTFGS-----1152
Db 1564 TLKHSDTQALQKTOQELQALQNLALAAASDNCPTIOEKETSADCVHPLEBKILLITBEL 1623
QY 1069 YFSTNFQEDA-----DLDQ-----DQDDSRQGNNSLDNQEAGLLKQKLA1L 1110
Db 1624 HQKTEQEKLLHEKNELEQAQVELKCEVHLMKSMIESLSLESLQHEKHEDTQQLLAD- 1682
QY 1111 LGNQFTQYYQNDKEI-----EFRIINVEKVE-----LSFRVEP-----1145
Db 1683 --KQOMQVVTQEBKELQOQTHEHLTAEVDHLKENIELGLNPKNEAQQKTTKEQCLLNENKE 1740
QY 1146 -----KLAKTLEDNGKTRIVLSDEMSLIIVNTTIETKTPMSAVPVPDTKW 1192
Db 1741 LEQSQRHLOCEIELMKSLLKXSALETLK-ESEQKVINLNQEMVMWLEMEELKNSQ-- 1797
QY 1193 EQYDPRTPLAAKTFVLKFKDQIPVDGSGNISDKWLASIPLVTHQOMRLSPVVKTIREL 1252
Db 1798 -----RTVIAE-----RDQL-----QDDLRESVEMSETQD 1823
QY 1253 GLKTEQOQQOQQOQQOQQO-----POKXAVRKEBEL-----ETYNPKDEFN1LNP 1297
Db 1824 DLKAQOALQOQKQVQOELTQSVLQEKISLLENQMLYNVATVKETLSERDDLN---Q 1879
QY 1298 TKAHRL-----TSLNVLNNDNYKIEDLVKIKNEAGDHQALAFSLRANKRLM-----NTP 1348
Db 1880 SKQHLSESETLS-LSLKEKEFALQEAQKADAAKTTDITEKISNIEBQLLOQATNLUK 1938
QY 1349 ITPADYNPPFYNE-----DWRSDIKYLNK-----GNVSSHQQQAAG-----GNQSGSLIQ 1395
Db 1939 ETLYERESLQCKEQLALNTEHLRETLKSKDALGKMEQERDEAANKVIALTEKMSLEE 1998
QY 1396 RLKNNI-----KBETP---TPALIALKORNTNLSNYSVKIIMIKPYL-----VERS 1440
Db 1999 QINENVTTLKEGEGKETFYLRPS---KQSSSQMEELRES---LTKDQLQLEBAEKE 2051
QY 1441 IGVPMSTGLDYGISQTDGTSSSSQKGFODFTQALGLKNTEY-----HGK----1489
Db 2052 I-----SEATHEIKMLTAKISSLEEEILQNASILNEAVSERENLHRSQQLV 2098
QY 1490 -----LGLSIRIPDPGNELAKIK--DASNK-----KGEKLL-KSYDLFKYNLNEYEKSP 1537
Db 2099 SELEQLSLTKSRDHAPQAKREKDEAVNKIASLAEIKILTKEMDEFDRDSKESLQEQSS 2158
QY 1538 KIAKGW-----TNTHPQOKEYPNQLKPE-----NYLNLVLNQPHKVTLYNSSDFI 1584
```

RESULT 31

```
US-09-949-016-10933
; Sequence 10933, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10933
; LENGTH: 3913
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10933
```

```
Query Match 2.5%; Score 238.5; DB 4; Length 3913;
Best Local Similarity 17.9%; Pred.No.18-06;
Matches 373; Conservative 286; Mismatches 690; Indels 729; Gaps 100;

QY 35 RGVNPTQGVISQLGLIDSVAFKPSTANFTSDYSQVKALLNGKTF---DPKSEFTDFVS 91
Db 1710 RSYDPSAGDVPQTQPEEPVSPKPS-----PTFMELEPKPTT-----1745

QY 92 KFDFTLNNGRVLEIPKYYQVVVISBSPDDKPKR-FRLGFHLKEKLEDGNIAQSATKFIY 150
Db 1746 -----SSIKEKVPQFMKAS--SEDDJHNRVLSKGMVRKEETH----ITTTTRMVY 1790

QY 151 LLP-----LDMPKAALGOYIVVDKFNFNLIHPLSNFSAQSISIKPALT 194
Db 1791 HSPPGGEGASERIEETMSVHDIMKAP--QSGRDPKPKELAGLPEH--KSAVSPDVHKSAE 1846

QY 195 RSSDFIAKLQNFNNQDE--LWVYLEKFFDLEALKANIR-----L 231
Db 1847 TSAQHAERKDNQMKPLERIEVHIEKGNQABTEVIRETKGHPKEMVYQKDLSRGDI 1906

QY 232 QTADSFKEGNLVDPFVYSFIRNPQNKQEWASD-----LNQDQKTVRLYLRTFSPQAK 285
Db 1907 NLKDLFPEKH---DAPPCSEEQGQEEBELTAESLPSLESSRVNTPVVSQEDSRPSSA 1963

QY 286 TILKDYKYKDETFLS--SI-----DLKASNGTSL-PANENDLKDQLDVLDLDD----VSDY 333
Db 1964 QIISDQSYKTLKLLSQHSIEYHDDSELSELRGSYFAEKMLLSEKLDVSHSDETSVTDH 2023

QY 334 FGQSQSETITSN---SQVKPVPASERSLKDRV-----KFKKQDQKPKRIEKFSLYEY 380
Db 2024 AGPPSSELQSGDKRSREKIATAPKKEILSKIYKVDSENGVGKVSQDE---HFDKVTVLHY 2080
```

Qy 381 DALSFYQLOELVSKPNSIKDLVNLARNLRF-----LGKYNFLPDDLASHLDYYFLV 435
Db 2081 SG-----NVSSPKH-----AMMRFTEDRLDRGREKLIYEDVRD----- 2115
Qy 436 SKAKIKOSSITKFLFIELPIKISLSSILGQEBNITLFEKEVTFKLDNFRDVEIEKAF 495
Db 2116 ----TVKEA-----EKLTEVSQFFRD-KTEK-- 2137
Qy 496 GLLYPGVNELEQARAKQASFEKEKKGLKPSQOKEENS-----KAINNOEGLEEDNI 552
Db 2138 -----LNDELQSPK-----KARPKNG-KEYSSQSPSSSEKVLLELLASNDWV 2183
Qy 553 TERL--PENSPIQOQENA-GLGASDPKPYMIKDVQORYYL-AKSOIELIKAKD----- 604
Db 2184 KARQHGPDGQGFPAEAKAPSLSPSEKMWLSQOTEDSKSTVEAKGSLQ-SKAPDGPQS 2242
Qy 605 --YTKLAKLGNR-----HTYNISRLKQLFDVNPRIIPSSRDIEKAKFVLDKTEKNK 655
Db 2243 GFOLKQSKLSIRLKFEGTHAKSKDMQSDRKSDGQSRIPVK----- 2286
Qy 656 YWQIYSSASPVQNKWSLFGYRYLLGLD-PKQTI--HELVKLGQAGLOFQEGYENLPS 711
Db 2287 ---IQESKLPVYQ---VFAREKQQAIDLDPDESVSQKDFMWLTK-DEHAQNEIIVN 2338
Qy 712 DPNLEDLKNIIRIKTFLPSQKDNFKLSLLDFNNYDGEIKAPEFGLPLPKELARNSSNS 771
Db 2339 DSGSDNVKQKTEMSSKAMPDSF-----SEQAKD--LACHITSDLA----- 2378
Qy 772 GGSQNSNPWQELIISQPKDQNLNQDLAQFSTKIWEKIIGDENEFQDNRLQVKKLD 831
Db 2379 -----TRGPWKKVF-----RTWES-SGATNNKSQEKLSHLVHD 2413
Qy 832 LQESWINKTRDNLWYTLGDKLVKPKNNLBAKFRQISNLQELITAFVTSAAALNNWNY 891
Db 2414 VRENHIGHP-----ESKSVQDN-----EFMSVTEREKL-----LTNGSL----- 2450
Qy 892 QDSGAKSTIIIEETAELDPKVEKGVADYOLKHYAIGFDNAGKFNQEVIRSSRTIY 951
Db 2451 -----ETKEMTVKSPK-----KVLRYEYVYVKEGHPGGLDQPSRRS----- 2488
Qy 952 LKTSKSKLEADTIDOLNAQVKNAPLGL--QSFYLDTERFVGVOFKLATSLAVOHKQKEKT 1009
Db 2489 -ESSAVSHI PVRVADERMWSNPIDGCEQASAPKHE-----LSQKLSQSSMSKETV 2540
Qy 1010 LPKLLNN-DGYTLIHDKLK-----PVIQISSPEK-----DWFEGLKNQNGQ 1052
Db 2541 ETQHENSIEDKVTYSEISKVSKHQSYVGLCPLEETETSTPKSPDSLEFSPG----- 2594
Qy 1053 SONVNVSTFGSIIESPYSTNFOEDADLDQGDSDSQNNSLDNOEAGLLKQKLAILLG 1112
Db 2595 -----ESP-----SSDVFHSPIDGLEKLAPLAQTEGGEIKTLPVYV- 2632
Qy 1113 NOFIOYQONKDEIEFEIINVEKVSFLRVEFKLAKTLEDNGKITRVLSDTMSLIYNT 1172
Db 2633 -SFVQVGQYKEKIO-----QGVKKIISQECKTVQ-----ETRGTFYTT 2671
Qy 1173 TIEK-----TPEMSAPVEV--FDTKWVQYDPTPLAAKTKFVLKFDQIPVDGSGNI 1223
Db 2672 RQKQPPSPQSGPEDTLEQVSLSDSGKSPITPETSSEVSY--EFTSKTP----- 2722
Qy 1224 SDKWLASIPLVTHQKRLSLPVVKTIRBELGLKTEOQQOQQOQQOQQOQKAVRKEELE 1283
Db 2723 -DSLAIYIP-----GKPSPIPEV-----SESEEEEOAKSTLSKQTTVEETAIVE 2765
Qy 1284 TYNPKDEFNINPLTKAHLRLTL-----SNLVNND-----PNYKIEDLVK-IKNE 1326
Db 2766 REMPNDVSKDSNQRKNNRVAIEPPPPPLDADQIESDKKHLYPEKEVDMEIYNLODE 2825
Qy 1327 AGDHQALAFSLRANNIKRLMNTPI--TFADYN-----PFFYVNEWRSIDKYLNN 1373
Db 2826 HDKYQLABEV-----IRVQPPSPVPGADVSDSDDESITYQPVVKKYTFKLKEVDDEQKE 2881
Qy 1374 KGNVS-----SHOQQAAGNGQS-----GLIQRLLNNIKPFTTFALLALKDRNNTNLSN 1423

Db 2882 KPXASAEKASQNELESNGSKDNFGLGLDSPQNE-----IAQNGNNDQSITE 2930
Qy 1424 YSKIMIKPKYLVERSIGVPMSTGLDGYIGSQDTGKTGSSSSSQ----- 1468
Db 2931 CS-----IATTAFFSHDTDATEIDSLDGY-DLOQEDDGLTESDKLPIQAMEIKKOIWT 2984
Qy 1469 ----KGFDDQFIQALGKNTYHKGKGLSLIRIPDGNELAKIKDASNGKEEKLKSYDL 1524
Db 2985 EGILKPADRSFQS-KLEVIEEGKV-----PDEKPPSKSSSEKTPDKT----- 3030
Qy 1525 FKYLNEYEKSPKIAKGTNIHDPDOKEYPNP--NOKLPENYLNVLNQPMKVTLYNSSD 1582
Db 3031 ----DOKSGAQPFLLGRHPRDSRVPDPTVFSYKVDSEFAT-----PEKTVATKGLD 3077
Qy 1583 FYNLFBPEBGRG----- 1597
Db 3078 F-----DPMNNRNGDEVPDSKSRDETKPFGLAVEDRSPATTPDTPPARTDESTPT 3131
Qy 1598 -----SGTKLQVIOQVNNVADMGSAVLTWY-----DKNIIIT 1632
Db 3132 SEPNNPFHEGKPFEMTRSGAIDMSKRDVFEERLQFQIHEHTSEKSGQGGEDKSMVT 3191
Qy 1633 NQPNVITANIADVFIKDVKELEDNTKLIAPNITQWENI-----SGSKERFPYK 1680
Db 3192 ATQO---PQSGDITVEINLERNVETPTVEPN---PSIPTSGECQBGTSSTSSGLEKSA 3243
Qy 1681 PTVFFGNWENESS-----MNSQAOPTTWEKIREG---FALQALKSS---FDQK 1723
Db 3244 AT-----NTSKVDPKLRTPIKMGISASITMTMKKEGPEITDKIEAVMTSCQGLENE 3294
Qy 1724 TRIFVL--TTNAPLPLWKYGPLGQNGENFKTQDWRLVFQDNDQIALRVQ----- 1773
Db 3295 TITMISNTANSQMGVRPHEKHDFOK-DNF-----NNNNLSDSTIQTDMNIMNIV 3343
Qy 1774 -----EODRPEKSSSEDK-----DKOK 1789
Db 3344 LTBHSAPTCTTEKDNPKVSVSSGKKTGVLQGHCVDRKQK 3381

RESULT 32

US-09-949-016-6978
; Sequence 6978, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6978
; LENGTH: 4377
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-6978

Query Match 2.5%; Score 238.5; DB 4; Length 4377;
Best Local Similarity 17.9%; Pred. No. 1.2e-06;
Matches 373; Conservative 286; Mismatches 690; Indels 729; Gaps 100;
Qy 35 RGVNTPQVISOGLGIDISVAFKPSIANFTSDYSQVKALLNGKTF---DPKSSSEFTDFVS 91
Db 2174 RSYDPSAGDVPTQPEPVPSPKPS-----PTFMELEPKPTT----- 2209

Db 1914 ERTOEQKTDIDRLQELLCASNRLQLEAEQQOIQBERELLRQKQKMAKAGPV---E 1970
Qy 1650 VKELEDNTKLIAPNITOWPNISSGKFKYKPTVFFGNWENENSSMNSQAQTPTWKIRE 1709
Db 1971 QOLLQETEKLM-----KEK-----LEVQQA---EKVRD 1996
Qy 1710 GFALQ--ALKSSFPQKTRTFVLTNAPLPLWKYGPLFGQNGPKTQDWRLVFNQDNDQI 1767
Db 1997 DLQKQVKALEIDVEQVSREI-----ELEQKNTLMDLRQONQALEKQL 2041
Qy 1768 AALR--VOEQ--DRP-EKSSSEKDKQKWKFKVVP 1798
Db 2042 EKWRKFLDEQAIDREHDFVQBEIQKLEQOLKVVP 2077

RESULT 36
US-09-538-092-1312
; Sequence 1312, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1312
; LENGTH: 2472
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/key: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q13813
US-09-538-092-1312

Query Match 2.4%; Score 229.5; DB 4; Length 2472;
Best Local Similarity 19.0%; Pred. No. 2.2e-06;
Matches 377; Conservative 287; Mismatches 618; Indels 699; Gaps 107;

Qy 182 NFSAQSIKPLALTRSSDFIAKLQFN-----QDELWVLEKFFDLEALKANIRLQ 232
Db 515 SLSAQBEK---ITALDEFATKLIQNNHYAMEDVATRDAL---LSRNALHERAMRRRAQ 568
Qy 233 TADFSPEKGNLDPFVYVFTIRNPQNKEMASDLNQDQKTVRLYLRTFSPQAKTILKDYK 292
Db 569 LAD-SFH-----LQQFPRDSDELKSW--VNEKMKTA-----TDEA 600
Qy 293 YKDETFL-SSIDLKASNGTSLPANEN--DLKDQLDVLDDVDSYFGQSETITNSQV-- 347
Db 601 YKDPNSLQGVKQHQAPAEALSANQSRIDALRAGQKLDVNHY--AKDEVAARMNEVIS 658
Qy 348 ---KPPVASE-RSLKDRVKFKDQKPRIBKFSLYEVDALS FYSQQLQELVSPNSIKDLV 403
Db 659 LWKKLLEATELKGILKEANQQQFNRNVEDIELWLY-----EVEGLASDDYKGLUT 711
Qy 404 NATLARNLRFSLGKYNFLPDDLASHLDY-----YFLVSKAKIKQSSITKK 448
Db 712 N---VQNLQ---KHALLEADVAAHQDRIDGVTIQAQRFQDAGHFDPAENIKKKQALVAR 765
Qy 449 LPIELPIKISLKSILGDQEPNKTIFKEVTFKLNFRDVELEKAF----- 495
Db 766 Y-----EALKPMPVARKQKLAADSLRLOQL-----FRDVEETWIREKEPIAASTNR 812
Qy 496 GLLYPGVNBELE--QARKAQARAFE---KEKSKKGLKFP--SQKEENS KAINNQGLEEDD 550
Db 813 GKDLIGVQNLKKHQALQAEIAGHEPRIKAVTQKG-----NAMV---EEGHFAAE 859

Qy 551 NITERLPENSPIQYOENAGLQASDPKPYMKDVQNRYYL-----AKSQIQE---LIKAK 603
Db 860 DVKAKLHE---LNQKWEALKAKASORRQDLESLOAQYQFADANEAEASWREKEPIVGST 916
Qy 604 DYTEK-----LAK---LLSNRHTYINISLR-LKEQ-----LF 629
Db 917 DYKGDSDSAEALLKHEALMSDL SAYSSIOALREQAQSCROQVAPTDDTGTGKELVALY 976
Qy 630 DVNPRIPSSRDIEKAKFVLDKTEKNKYWOIYSSASVPFQNKWSL-----FCYYRYLLG 682
Db 977 DYQEKSPREVTWKKGDIILTLNSTKDW-----WKVEVNDRQGFVPAAYVKK 1023
Qy 683 LDPKQT-----IHELVLKLGK- 698
Db 1024 LDPAQASRENLEEQSIALRQEQIDNQTTRITKEAGSVSLRMKQVVEELVHSLLEGEKR 1083
Qy 699 AGL-----QFEGYEN-----LPSDFNLEDLKNIRIKTPLFS--QKDNFKL 736
Db 1084 KGMLEKSCCKFMLFREANELQOWINEKEAALTSEEVGADLEQVEVLQKKFDDFKD-LKA 1142
Qy 737 S---LILDFNNYYDG-----ETKAPB---FGLPLPLPKELRRNNSNGSSQNSNSP 780
Db 1143 NESRLKDINKVAEDLESEGLMAEEVOAQVQQEYVGM-----MPRDETD--KTASP 1191
Qy 781 WEQ-----ETISQFKQNLSPQDLAQF-----STKIW--- 808
Db 1192 WKSARLMVHTVATFNSIKELNERWRSLOQLABERSQLIGSAHEVQRFHRDADETKEWIEE 1251
Qy 809 -----EKITGDENE?-----DONNRL-----QYKLLKDLO 833
Db 1252 KNQALNTDNYGHDLASVQALQRKHGFERDLAALGDKVNSLGETABRLTQSHPESAEDLQ 1311
Qy 834 ESWINKTRDNLVWTVGLDKLVKPKNNLEAKFRQISNLQELLTAFVTSAAALSNNWNYQD 893
Db 1312 EKC---TELNQAWSSLG-----KRADQKAKLGDSDHDLORFLSDFRDLMSWIN- 1356
Qy 894 SGAKSTIIPETAEALDPKVKKEKGVADVYQLRPH--YAIGFDDNAG-----KFNQEVIRS 945
Db 1357 -GIRGLVSSDELA-----KDVTAEGAL-LBRHQEHRTEIDARAGTFOAFEFGQQLLAH 1408
Qy 946 SRTTYLTKSGSKLEADTIDQLNQAKVA-----PLGLQSF----- 982
Db 1409 GH---YASPEIKOKL--DILDQERADLEKAVQVORRMMLDQCLEQLFHRDCEQAENWMAA 1463
Qy 983 ---YLDTFRFGVFKLATS LAVQHK-----QKEK-----TLPKKLNNDGVTLIHD-- 1024
Db 1464 REAFLNTEKGDSDLSVEALIKKHEDFDKAINVQBEKTAALQAFADQLIAAGHYAKGDIS 1523
Qy 1025 -----KLKXPVPIQISSPEKD-----WFEGKLNQNGOSQNVNVS 1059
Db 1524 SRNREVLDORWRRLKAQMIEKRSKLGESQTLQFSDRVDBIEAWISEKL-QTASDESYPK 1582
Qy 1060 TFGSIIESPY-----FSTNFOEDADLDQDQDSDSQGNNSLD-----NOEAGLLKOKLA 1108
Db 1583 T---NIQSKHQKHQAPEALHANADRIRGV1D--MGNSLIBERGACAGNEDA--VKARUA 1634
Qy 1109 ILLGN-QFI-----QYQONDKIEPEIINVEKVSLSF---RVEPKLAKTLEDNGK 1156
Db 1635 ALADQWQFLVQKSAEKSQKLEANKQNFN-----TGIKDIAFWLSEVEALLAS--EDYCK 1688
Qy 1157 TIRVLSD-----ETWSLIVNTTIEKTPMSA-----VPEVFDTKWVE-----QYDP 1197
Db 1689 DLASVNNLLKKHQLLEADISAHEDRLKDLNSQADSLMTSSAFDTSQVKDKRDTINGFQK 1748
Qy 1198 RPLAAKTKFVLKFKDQIPVDGSGNISDKWLASIPLVIIHQMLRSLSPVVKTTRELGLKTE 1257
Db 1749 IKMSAASRRAKL-----NESHK-----LHQFFRDMDBEESWIKKLLVG 1788
Qy 1258 QOQOQOQOQOQOQOQKAVRKEEELTYNPKDEFNLPNTKAHRLTSLNVLNNDPNYK- 1316
Db 1789 SEDYGRDLTGVLNLRKKHKLLEAELAAHEPA-----IQGVLDTGKKLSDDNTTIGKEIQOR 1844

QY 1317 ----IEDLKVINERAGDHOIAFSLRANNIKELMNTPTITADYNPPFFYYNEDWRSIDKXN 1372
DB 1845 LAQFVHWEKELK-----OLA-AARGORLESL-----EYQOFVANVEE---EEAWIN 1887
QY 1373 NKGNVSSHQQAAGNQGSL--TORLNKNIKPETFTPALIALKDRNNTNLSYSDKIIM 1430
DB 1888 EKWTL-----VASEDYGDTLAAIQGLK--KHEAFETAFTVHKDRVNDVCTNGQD--L 1936
QY 1431 IKPKYLVERSIGVPWSTGLDGYI-----GSBQTDGTSSSSQ--QKGFQDQFIQALGLKN 1483
DB 1937 IKKNHHEENISSKMK-GLINGKVSDELEKAAQARKANVDENSAFLQFNWKAADVES----- 1990
QY 1484 TEYHGKLGSLIRIPDPGNELAKIKDASNKGBEKLKSYDLFKVNLNEYEKSPKIAGW 1543
DB 1991 --WIGEKENSLKTDYGRDLSSV-----QTLTKQETFDAGLQFQO-----EGI 2033
QY 1544 TNI-----HPDQKEYPNPNOKLPENYLNLVLPQWVTLVYNSSDFITNLVPEPE 1592
DB 2034 ANITALKDQLLAAKHVQSKAEARHASLMKWSQLAN----- 2071
QY 1593 GSDRGSGTKLKQVIOKQVNNVADWGSAYLTF-----WYDKNIITNQPNVITANIA 1643
DB 2072 -----SAARKKKLEAQ--SHPRKVEDLFTPAKKASAFNWE-NAEEDLTPVRCNSL 2123
QY 1644 DVFIKDVKELED-----NTKLIAPNITQW-----WPN 1671
DB 2124 BE-IKALREAHDAFRSSLSAQAQFNQLAELDRQIKSPRVASNPYTWFTMEALBETWRNL 2182
QY 1672 SGSSEKFPKPTVFFGNWENENSSMNSQAQTPTEKIREGFALQALKSFD---QKTRTFV 1728
DB 2193 ----QKLIK-----BRELEQEQRRQEBENDKLROEFAQHA--NAFHQWQIETRYL 2228
QY 1729 L 1729
DB 2229 L 2229

RESULT 37

US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TRANSCRIPTION: TRANSCRIPT-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET B.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: HUMAN
US-08-353-700-1

Query Match 2.3%, Score 228; DB 1; Length 3248;
Best Local Similarity 18.1%; Pred. No. 4.1e-06;
Matches 376; Conservative 346; Mismatches 774; Indels 578; Gaps 93;

QY 61 NPTSDYSQVVKALLNGKTFDPKSSSEFTDFVSKPDEL-TNNGRTVLTEIPKKYQVVISFSP 119
DB 1066 NRKNELEQLKEAF-----AKEHQEFUTKLAFABERNQNLMELETVOQALRSEWTD 1116
QY 120 EDDKERFRLGPHLKEKLEDGNIASQATFYLLPDMPKAALGQSYIVDKNFNNLI--- 176
DB 1117 NON-----NSKSEAGGLKQE---IMTLKEQNK-----MQKEVNDLLQEN 1153
QY 177 -----IHPISNFAQSIKPLALTRSSDPIAKLQ---FNNODELWV-----YLEK 218
DB 1154 EQLMKVMKTKHECONLESEPIRNSVKERESE---RNQC�FKPQMDLEVKELISLDSYNAQ 1209
QY 219 FFDLEALKAN-----IRLOTADFSFEKGNLVD-----PFV 248
DB 1210 LVQLEAMLRNKLKLOEKEKECEQLQHELOTIRGDLTSLNQDMQSEISGLKDCEDAE 1269
QY 249 YSFTIRNPQNKQEWASDLNQDKTVRLYLRTFFS-----POAKTILKDYKYKDETFLSSID 303
DB 1270 EKYISGPH-----LSTQSDNAHLQCSLQTTMKNLEKEICEILOAEKYELVTELNDSR 1325
QY 304 LKASNGTSLFANE-----NDLKQDLVDLVDVDFGQ--SETTITSNQVKEPVP 351
DB 1326 SECITATRKMAEEVGKLLNEVKILNDDSGLIHGL--VEDIPGEGFGEQNPQHPVSLAP 1383
QY 352 ASERSLKDRVKFKDQKQPRI-----EKFSLYEYDALSPYQOLQELVSKPNSIKDLVNATL 407
DB 1384 LDESNSYEHLTSDKEVQMHFAELOEKLPSIQSHKILHDQHCMSKMSQLQYVDSLK 1443
QY 408 ARNLRFSLGKYNF---LFDOLASHLDYFL--VSKAKIKQSSITKCL-----FIELP 454
DB 1444 AENLVLTNLNFOGDLVKEMQLGEBGLVPSLSSCVPDSSSLSSGLSDSYFALLBQT 1503
QY 455 IKISLKSILGDBPN---IKTLFEKVTFKLDFNPRDVEIEKAFGL-----LYPG 501
DB 1504 GDMSSLNLEGAVSANQCSVDVFCSSLQTYVDSLKAENLVLTNLNRFQGDVLKEMQLG 1563
QY 502 VNEELEQARKAQRASFEKESKKGKLFESQ--OKEENSKAINNOEGLE----- 547
DB 1564 LEEGLVPSLSSCVPDSSSLSSGLSDSYFALLBQTGDMSSLNLEGVVSANQCSVDVDF 1623
QY 548 ----EDDNITERLPENSPIOYQENAGLGAEPDPYMIKDVQNGRYIYLAQSIQELIKAK 603
DB 1624 CSSLQENLTKETPSAPAKGVESLES---CEVYR---QSLKLEKESQGIKMKK 1675
QY 604 DYTAKLALLSNRHTYNIISRLKEQLFDVNPRIPSRDIEKAKFVLDKTEKNKYWQIYSSA 663
DB 1676 BIQLEQLLSERQELDCLR-KQYL-----SENEQWQOKLTSV 1712
QY 664 SPVQNKWSLFGYYRYLLGLDPKQTIHELVLKQAGLOFEGYENLPDFNLEDLKNRI 723
DB 1713 TLEMESK-----LAAEKQTEQLSLEL-EVARLQLOGL-----DLSSRLIGDIT 1756
QY 724 KTLPLFSOKDNFKLSLLDFNNYDGEI--KAEFGFLPLFLPKELRRN----- 767
DB 1757 EDAIQGNES-----CDISKEHTSETTERTPKHDVHOICDKDAQDNLNDIEKITGTGAV 1811
QY 768 --SSNSGGSQNSNPWE-----QEIIISQ--FKDON-----LSNQDLAQFSTK 806
DB 1812 KPTGCSGEGQPDNTNYPPEGDKTQGSSECSISLFSFGPNALVPMDFLGNQEDIHNLQLR 1871

Qy 249 YSFIRNPONQKEMASDLNODQKTVRLYLRTFFS-----PQAKTILKDYKYKDYKDTFLSID 303
 Db 1270 EKYISGPHE-----LSTQNDNAHLQCSLQTTMKNLELEKICEILOAKYELVTELDSR 1325
 Qy 304 LKASNGTSLPANE-----NDLKQDLVDLLDVSDVFGQO-SETITSNSOVKVPV 351
 Db 1326 SECITATRKMAEVOGKLLNEVKILNDDSGLLHGL--VEDIPGGEFGEQNEQHPVSLAP 1383
 Qy 352 ASERSLKDRVFKKQDQKPRI-----EKFSLYEYDALSFYSQLBELVSKPNSIKDVLNATL 407
 Db 1384 LDENSIEYHLTLDSEKQVHFAELQEFSLQSEHKILHDQCMQSSKMSSELQTVVDSLK 1443
 Qy 408 ARNRPFSLGKYNF--LPDDLAHLDDYFL--VSKAKIKQSSITKYL-----FIEPL 454
 Db 1444 AENLVLTNLRNFQDLVKEMQLGLEGLVPSLSSSCVPDSSSLSSLDGDSFYRALLEQT 1503
 Qy 455 IKISLKSSILGDOBNP--IKTLPEKVTFFKLDNPRDVEIEKAFGL-----LYPG 501
 Db 1504 GMSLSLSNLEGAVSANQCSVDVFCSSLQTVVDSLKAENLVLSLNRNLFQDLVKEMQLG 1563
 Qy 502 VNEELEQARKQARASFEKSKGKLFQSO--QKEENSKAINNOEGLE----- 547
 Db 1564 LEEGLVPSLSSSCVPDSSSLSSLGDSFYRALLEQTGMSLSLSNLEGVVSANQCSVDV 1623
 Qy 548 ----EDDNIATERLPNSPIYOQOENAGLSPDKPMTKDVQNRYYLAKSQIQLIKAK 603
 Db 1624 CSSLQENLRTKETPSAPAKGVEELES-----CEVYR----QSLKLEERMESQIGMKN 1675
 Qy 604 DYTAKLAKLSNPHYTNISLRKEQLFDVNPRISSRDIEKAKFVLDKTKKMYQLYSSA 663
 Db 1676 EFOLEQLLSSERQBEDCLR-KOYL-----SENQWQOKLTSV 1712
 Qy 664 SPVFQNKWSLFGYYRLLGLDPKQTHIELVKLGKAGLQFEGYENLPSPDFNLEDLKNIRI 723
 Db 1713 TLEMESK-----LAABKKQTEQLSLEL-EVARLQQL-----DLSSRSLGIDT 1756
 Qy 724 KPTLFSQKDNFKLSLDFNNYVDGEI--KAPFGLPLPLPKELRN----- 767
 Db 1757 EDIAIQORNES-----CDISKEHTSETTPKHVDHQICDKAQOQDLNLDIEKITETGAV 1811
 Qy 768 ----SSNSGSGSONSPWE-----QELISO--FKDQO-----LSNODQLAQFSTK 806
 Db 1812 KPTGCSGQSPDTNYPPEGDKTQGSSEICISLSPGNALVPMDFLGNQEDIHNLQR 1871
 Qy 807 IWEKIIGDENBFDQNNRLQYLLKDLQ-----ESWINKTRD-----NLYWYTLGDKLV-- 855
 Db 1872 VKE-----TSNENLRLLH-VIEDRDRKVESLLNEMKELDSKLHLQEVQLMTKIEACIE 1923
 Qy 856 -----KPKNNLEAKFRQIS-NLQELLTAPYTSALSNWNYYQDSGAKSTIIFREI 905
 Db 1924 LEKIVGELKKENSIDLSEKLEYFSCDHOELLQRTVETSEGLNSDLEMHADKSSRED-IGDNV 1982
 Qy 906 AELDPKVKKEK--VGADVYQLAF-----HYAIGFD-----DNAGKFNQE 941
 Db 1983 AKVNSWKRERFDVNELSRISSEKASIEHAYLEADLEVUVQTEKLCLEKDNENK--QK 2040
 Qy 942 VIRSSRTIYLTGSKSL--EADT-----IDLQNAVKNAPLGLQSFYLDTRFPGVF 992
 Db 2041 VIVCLEBELSVVTSERNQRLGELDTMSKKTALDQLSEKWKETQBELESHQSEC----- 2094
 Qy 993 QKLATSLAVOHQKQKKT-IPKLLNNDGYTLIHDKLKKVPVPOISSPEKD-----W 1042
 Db 2095 --LHCIVQAEAEVKEKTELLQTLSSDVSELKDKTH--LOEKLOSLEKDSOALSITKE 2149
 Qy 1043 FEGLKQNGOSQNVV-----STFGSIESPYSTNFQO--DADLDQD-----QDD 1087
 Db 2150 LENQIAQLAKKEKELLVKESESQARLSESDYEKLVNSKALEALVEKGFALRLSTQEE 2209
 Qy 1088 SRQGNNSLNQEAAGL-LKQKLAILLGNQFIQYQQNDKEIFEIINVEKVSLSRVEPK 1146
 Db 2210 VHQLRGIEKLRVRIEADKQLHIAEKLEREREND-SLKDKVENLERELQMS----- 2262

Qy 1147 LAKTLDNGKTIIRVLSDETMSLIIVNTTIEKTPMSAVPEVFDTKWVEQYDPRTPLAATK 1206
 Db 2263 -----FENQELV--ILDAENSKAEVELTKQIEBMASSLSKIFELDLVLRSEKENLTKQIQ 2316
 Qy 1207 FVLKFKDQIPVDGSGNIS--DKWLASIPLVHOOMLRLSPVVTIIRBELGLKTKTEQOQOQOQ 1264
 Db 2317 -----EKQQLSELDKLLSSFPKSLLE-----KEQABIQIKESKTAVEM 2356
 Qy 1265 QOQOQOQOKKAVRK-----EEL-----ETNYP--KDFENILNPLTKAHLRLTSLNVLNNDPN 1314
 Db 2357 LQNLKELNEAVALCQDQOIMKATQOSLOPPPEEHEHOLRNSIEK-----LRARLEADEK 2411
 Qy 1315 YKTEDLKVIKNEAGDHQALFSLRANNIKRLMNTPIITFADYNPPFFYNNEDRSIDKYLNNK 1374
 Db 2412 KQLCVLOQLK-ESEHADLLKGRVENLERELEIART-----NOEHAALAE-NSK 2459
 Qy 1375 GNVSSHQOQOAGNQO-----SGLIQLRLNK-----NKPETFTAL- 1410
 Db 2460 GEVETLKAKIEGTMQSLRGLDELVDVTIRSEKENLTNLSQEKERISELEINSFENILQ 2519
 Qy 1411 -----TALKDRNT-----NLSNVSDKIIMKP 1433
 Db 2520 EKEQEKVOMKEKSTAMEMLOTLKELNERVAALHNDQEAQKAEQNLSSQVECLELEKA 2579
 Qy 1434 KYLVERSIGVPMSTGLDYGIGSEQTKDGTSSSSQKQFDQ---DFIQALGLKNTYHGKL 1490
 Db 2580 QLL-----QGLD-----EAKNNYIVLQSSVKGLIOEVEDGKQKLEKDEIS-- 2621
 Qy 1491 GLSIRIFDPGNELAKIKDASNKKEEKLKSYDL-FKYNLYNEYKKSPPKAKGWTNHPD 1549
 Db 2622 RLKNOIQDQEOQLVSKL---SQVEGEHQLWKEQNLLEENLTVELRQKIQVLQSKNASLQ-D 2677
 Qy 1550 QKEYPNFNQKLPENYLNVLNQPWKVLYNESDPITNLVEPEGSDRGSGTKLQVIOKQ 1609
 Db 2678 TLEVLOSSYNKLENELEL-----TMDQMS--FVEKVNMTAKETELQREHEM 2724
 Qy 1610 VNNYADWGSAYLTFWYDKNIITNQPNVITANIAADVFIKDVKVELEDNTKLIAPNITQWPP 1669
 Db 2725 AQK-----TAELOE-----ELSGEKNRLAGELQLLE 2751
 Qy 1670 NISGSKFKYKPTVFFGNWENENSMNS-----QAQPTWEKIREGFALQALK-SSPDQ 1722
 Db 2752 EIKSKDQLKELTL-----ENSELKSLDCMHKQDQVEKGRVREEIAEYQLRLHEABK 2804
 Qy 1723 KTRFVLTTNAPLPLWKYVGLFGQGNFKTODWRLVFQNDNOIALARVQEQDRPEKSS 1782
 Db 2805 KHQALLDQTNQYEV-----EIQTRKELKTSKEECLSSQKL-EIDLKSKS 2849
 Qy 1783 BDKDKQKWPKVPIBEMFNS--GNIRFVEMQ 1814
 Db 2850 EELNNS--LKATTQILBELKTKMDNLKYVNLQK 2881

RESULT 39
 US-09-134-001C-5178
 ; Sequence 5178, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 5178
 ; LENGTH: 1010
 ; TYPE: PRN
 ; ORGANISM: Staphylococcus epidermidis

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2005, 20:53:38 ; Search time 243 Seconds
(without alignments)
3959.658 Million cell updates/sec

Title: US-10-607-631-8
Perfect score: 9732
Sequence: 1 MKNKSTLLATAAAIIGST.....TNNAFNVFKEPNISKIVE 1879

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 03.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9701	99.7	1879	Q71A34	Q71A34 mycoplasma
2	9435.5	97.0	1878	Q71A35	Q71A35 mycoplasma
3	9402	96.6	1871	Q719M1	Q719M1 mycoplasma
4	2816	28.9	560	Q9KX9	Q9KX9 mycoplasma
5	2770	28.5	560	Q9KX7	Q9KX7 mycoplasma
6	523	5.4	1405	Q8KWK1	Q8KWK1 mycoplasma
7	502.5	5.2	1427	Q8G122	Q8G122 mycoplasma
8	473.5	4.9	1296	Q9KX3	Q9KX3 mycoplasma
9	410.5	4.2	1078	Q6R5H9	Q6R5H9 mycoplasma
10	407	4.2	1037	Q6R518	Q6R518 mycoplasma
11	405	4.2	5767	Q8I525	Q8I525 plasmidium
12	404	4.2	1063	Q6R510	Q6R510 mycoplasma
13	399	4.1	1077	Q6R514	Q6R514 mycoplasma
14	395.5	4.1	2965	Q8WR55	Q8WR55 mycoplasma
15	394	4.0	1099	Q49560	Q49560 mycoplasma
16	393	4.0	1092	Q07132	Q07132 mycoplasma
17	391.5	4.0	1092	Q6R517	Q6R517 mycoplasma
18	391.5	4.0	1108	Q49542	Q49542 mycoplasma
19	390.5	4.0	1098	Q6R516	Q6R516 mycoplasma
20	390	4.0	2740	Q7RF82	Q7RF82 plasmidium
21	390	4.0	2976	Q8WR54	Q8WR54 plasmidium
22	389.5	4.0	2664	Q7RE10	Q7RE10 plasmidium
23	389	4.0	1093	Q6R5H8	Q6R5H8 mycoplasma
24	387.5	4.0	2770	Q7YU99	Q7YU99 plasmidium
25	387.5	4.0	2969	Q7KFP73	Q7KFP73 plasmidium
26	385.5	4.0	2261	Q7RRR9	Q7RRR9 plasmidium
27	384	3.9	2757	Q7RRR9	Q7RRR9 plasmidium
28	382.5	3.9	1098	Q6R512	Q6R512 mycoplasma
29	380.5	3.9	2977	Q8WP99	Q8WP99 mycoplasma
30	380	3.9	2747	Q8BJX9	Q8BJX9 plasmidium
31	378.5	3.9	3427	Q6YA79	Q6YA79 plasmidium

ALIGNMENTS

RESULT 1

Q71A34 Q71A34 PRELIMINARY; PRT; 1879 AA.
AC Q71A34;
DT 05-JUL-2004 (TremBLrel. 27, Created)
DT 05-JUL-2004 (TremBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TremBLrel. 27, Last annotation update)
DE F216 surface protein.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=232;
RA Wilton J., Cordwell S., Minion C., Hahn M., Walker M., Djordjevic S.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF540381; AAQ11196.1;
SQ SEQUENCE 1879 AA; 215648 MW; 6F255658D97BB8E2 CRC64;

Query Match 99.7%; Score 9701; DB 2; Length 1879;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1875; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	1	MKNKSTLLATAAAIIGSTVFVTGVLASVKYRGVNPQTQGVISQLGIDSVAFKPSIA	60
Db	1	MKNKSTLLATAAAIIGSTVFVTGVLASVKYRGVNPQTQGVISQLGIDSVAFKPSIA	60
QY	61	NFTSDYQSVKXALLNGKTFDPKSSBFTDPVAFKDFLTNNGRVTLEIPKKYQVVISFSPB	120
Db	61	NFTSDYQSVKXALLNGKTFDPKSSBFTDPVAFKDFLTNNGRVTLEIPKKYQVVISFSPB	120
QY	121	DKERFRLGFLHKEKLEGNIAQSAKFTYLPLOMPKALGOYSYIVDKNFNNLIHPL	180
Db	121	DKERFRLGFLHKEKLEGNIAQSAKFTYLPLOMPKALGOYSYIVDKNFNNLIHPL	180
QY	181	SNFSAQSIKPLATRSSDFIAKLNQNNODLWVLEKFFDLKALKANRLQTADFSFEK	240
Db	181	SNFSAQSIKPLATRSSDFIAKLNQNNODLWVLEKFFDLKALKANRLQTADFSFEK	240
QY	241	GNLVDPFVYSFIRNPQNKWASDLNQDKTVRLYLRTFSPOAKTILDKYKDTFTLS	300
Db	241	GNLVDPFVYSFIRNPQNKWASDLNQDKTVRLYLRTFSPOAKTILDKYKDTFTLS	300
QY	301	SIDLKASNGTSLPANENDLKQDLVDLLVDVYFGQSETITSNSQVKKVPASERSLKDR	360
Db	301	SIDLKASNGTSLPANENDLKQDLVDLLVDVYFGQSETITSNSQVKKVPASERSLKDR	360
QY	361	VFKKQKQPRIEKFSLEYDALSFYSQQLVSKENSIKDLVNATLARNRFSLGKYNF	420
Db	361	VFKKQKQPRIEKFSLEYDALSFYSQQLVSKENSIKDLVNATLARNRFSLGKYNF	420
QY	421	LPDDLASHLDYFVLVSKAKIKQSSITKKLFIELPIKISLKSILGQBPNIKTLPKEVT	480
Db	421	LPDDLASHLDYFVLVSKAKIKQSSITKKLFIELPIKISLKSILGQBPNIKTLPKEVT	480

```
QY 481 PKLDNFRDVEIKAGLLYPGVNELEOARKAQRASFEKESKGLKBFSSQOKEENSKAI 540
DB 481 PKLDNFRDVEIKAGLLYPGVNELEOARKAQRASFEKESKGLKBFSSQOKEENSKAI 540
QY 541 NNQGLEBDDNITERLPENSPITQYOENAGLGPSPKPYMIKDVQONQRYLLAKSQIOBELI 600
DB 541 NNQGLEBDDNITERLPENSPITQYOENAGLGPSPKPYMIKDVQONQRYLLAKSQIOBELI 600
QY 601 KAKDYTKLAKLLSNRHTYNIISRLKEQLFDVNPRIIPSSRDTEKAKFVLDTKCKNKYQIY 660
DB 601 KAKDYTKLAKLLSNRHTYNIISRLKEQLFDVNPRIIPSSRDTEKAKFVLDTKCKNKYQIY 660
QY 661 SSASPVFNQKSLFGYRYLLGLDLPKQTHIELVKGQKAGLQFEGYENLPDSFNLDELKN 720
DB 661 SSASPVFNQKSLFGYRYLLGLDLPKQTHIELVKGQKAGLQFEGYENLPDSFNLDELKN 720
QY 721 TRIKTPFSQDNFKLSLLDFNNYDGEIKAPEFGLPLFLPKELARRNSSSGGQNSNSP 780
DB 721 TRIKTPFSQDNFKLSLLDFNNYDGEIKAPEFGLPLFLPKELARRNSSSGGQNSNSP 780
QY 781 WEQETISQFQDNLSNQDLAQFSTKIWEKIIGDENEFQNNRLQYKLLKDLQESWINKT 840
DB 781 WEQETISQFQDNLSNQDLAQFSTKIWEKIIGDENEFQNNRLQYKLLKDLQESWINKT 840
QY 841 RDNLTYTYLGDKLKVKPKNNLEAKPRQISNLQELLTAFYTTSAALSNWNYYQDSGAKSTI 900
DB 841 RDNLTYTYLGDKLKVKPKNNLEAKPRQISNLQELLTAFYTTSAALSNWNYYQDSGAKSTI 900
QY 901 IPEETAELOPKVEKVGADVQLKPHYAIGPDDNAGKQENQEVIRSSSTIYLYKTSKSKL 960
DB 901 IPEETAELOPKVEKVGADVQLKPHYAIGPDDNAGKQENQEVIRSSSTIYLYKTSKSKL 960
QY 961 EADTTLDQNLQAVKNAPLGLQSYLDTERPFGVQKLAATSLAVQHKQKXETLPKLLNNQGYT 1020
DB 961 EADTTLDQNLQAVKNAPLGLQSYLDTERPFGVQKLAATSLAVQHKQKXETLPKLLNNQGYT 1020
QY 1021 LIHDKLKKPVPPIQISSPEKQWFEKLNQNGSQNVNVTFGSIIIESPYFTNFQEDADL 1080
DB 1021 LIHDKLKKPVPPIQISSPEKQWFEKLNQNGSQNVNVTFGSIIIESPYFTNFQEDADL 1080
QY 1081 DQGGDDSRQNSLNDQAGLLKQKLAILLGNQFIQYQONDKEIEBEIINVEKVSLS 1140
DB 1081 DQGGDDSRQNSLNDQAGLLKQKLAILLGNQFIQYQONDKEIEBEIINVEKVSLS 1140
QY 1141 PRVEPKLAKTLEDNCKTIRVLSDETMSLIIVNTTEKTPMSAVPEVPTKWVQYDPRTP 1200
DB 1141 PRVEPKLAKTLEDNCKTIRVLSDETMSLIIVNTTEKTPMSAVPEVPTKWVQYDPRTP 1200
QY 1201 LAAKTKFVLKFKDQIPVDGSGNISDKWLASIPLVHQQMLRLSPVVKTIRELGLKTEQQQ 1260
DB 1201 LAAKTKFVLKFKDQIPVDGSGNISDKWLASIPLVHQQMLRLSPVVKTIRELGLKTEQQQ 1260
QY 1261 QOQOQOQOQOQKAVRKEELETYNPKDEFNINPLTKAHLRLTSLNVLNNDPNYKIEDL 1320
DB 1261 QOQOQOQOQOQKAVRKEELETYNPKDEFNINPLTKAHLRLTSLNVLNNDPNYKIEDL 1320
QY 1321 KVINEAGDHQALFSLRANNIKRLMNTPIITADYNPFYFNEDWRSIDKYLNNKGNVSSH 1380
DB 1321 KVINEAGDHQALFSLRANNIKRLMNTPIITADYNPFYFNEDWRSIDKYLNNKGNVSSH 1380
QY 1381 QOQAAAGGQSGSLIQLRLNKNIKPETFTPALIALKDRNNTNISYSDKIIMIKPYLVERS 1440
DB 1381 QOQAAAGGQSGSLIQLRLNKNIKPETFTPALIALKDRNNTNISYSDKIIMIKPYLVERS 1440
QY 1441 IGVPHSTGLDGYIGSEQTKDGTSSSQOQGFDDFIQALGLKNTYEHGKGLSLTRIFDPG 1500
DB 1441 IGVPHSTGLDGYIGSEQTKDGTSSSQOQGFDDFIQALGLKNTYEHGKGLSLTRIFDPG 1500
QY 1501 NELAKIKDASNNKGBEKLKSYDLFKNYLNEYEKSKPKIAKGWNIHPDQKEYPNPQKL 1560
DB 1501 NELAKIKDASNNKGBEKLKSYDLFKNYLNEYEKSKPKIAKGWNIHPDQKEYPNPQKL 1560
```

```
QY 1561 PENYLNILVLPQWKVTLNYSDFITNLFVPEPESDRSGTGKTKQVIOQVNNNNYADWGS 1620
DB 1561 PENYLNILVLPQWKVTLNYSDFITNLFVPEPESDRSGTGKTKQVIOQVNNNNYADWGS 1620
QY 1621 YLTFWYDKNIIITQPNVITANIADVFIKOVKELEDNTKLIAPNITQWPNISGSKFVK 1680
DB 1621 YLTFWYDKNIIITQPNVITANIADVFIKOVKELEDNTKLIAPNITQWPNISGSKFVK 1680
QY 1681 PTVPFQGNWENSSMSQAOTPTWEKIREGFALQALKSSFDQKTRTFVLTNAPLPLWKY 1740
DB 1681 PTVPFQGNWENSSMSQAOTPTWEKIREGFALQALKSSFDQKTRTFVLTNAPLPLWKY 1740
QY 1741 GPLFGQGNPFKTDQWRLVFNQDNDQIAALRVQEQDRPEKSSDKQKWKFKVPIPEE 1800
DB 1741 GPLFGQGNPFKTDQWRLVFNQDNDQIAALRVQEQDRPEKSSDKQKWKFKVPIPEE 1800
QY 1801 MFNSGNIRFVGVMQIQGPNLTLVPVINSVIYDFYRGTDGSDNVANLNVAPWQVKTIAFT 1860
DB 1801 MFNSGNIRFVGVMQIQGPNLTLVPVINSVIYDFYRGTDGSDNVANLNVAPWQVKTIAFT 1860
QY 1861 NNAFNNVFKEFNISKKIVE 1879
DB 1861 NNAFNNVFKEFNISKKIVE 1879

RESULT 2
Q71A35 PRELIMINARY; PRT; 1878 AA.
AC Q71A35, 2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE P216 surface protein.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OC NCBI_taxid=2099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J;
RA Wilton J., Cordwell S., Minion C., Hahn M., Walker M., Djordjevic S.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF540380; AAQ11195.1; -.
SQ SEQUENCE 1878 AA; 215617 MW; B6FDA7B8B2DB9037 CRC64;

Query Match 97.0%; Score 9435.5; DB 2; Length 1878;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1827; Conservative 17; Mismatches 29; Indels 11; Gaps 3;

QY 1 MOKKSTLLATAAAIIGSTVGTGVGLASKVKYRGVNPQTGGVISQLGLDSVAFKPSIA 60
DB 1 MOKKSTLLATAAAIIGSTVGTGVGLASKVKYRGVNPQTGGVISQLGLDSVAFKPSIA 60
QY 61 NFTSDYQSVKALLNGKTFDPKSEFTDPVSKPFDLTNNGRTVLEIPKKYQVWISFSE 120
DB 61 NFTSDYQSVKALLNGKTFDPKSEFTDPVSKPFDLTNNGRTVLEIPKKYQVWISFSE 120
QY 121 DDKERFRLGPHLKEKLEDGNIAGSATKFIYLLPLDMPKAALQGVSYIVDKNFNNLIHPL 180
DB 121 DDKERFRLGPHLKEKLEDGNIAGSATKFIYLLPLDMPKAALQGVSYIVDKNFNNLIHPL 180
QY 181 SNFSAQSIKPLALTRSSDPIAKLNQFNQDELWVYLEKPFDEALKANIRLOTADFSFEK 240
DB 181 SNFSAQSIKPLALTRSSDPIAKLNQFNQDELWVYLEKPFDEALKANIRLOTADFSFEK 240
QY 241 GNLVDPPIVYSFIRNPQKQEWASDLNQDKTQVRLYLRTFSPQAKTILKDYKXDETFLS 300
DB 241 GNLVDPPIVYSFIRNPQKQEWASDLNQDKTQVRLYLRTFSPQAKTILKDYKXDETFLS 300
QY 301 SIDLKASNGTSLFANENDLKQDLVDLLDVSDFGQSGSETITSQVQKVPASERSLKDR 360
DB 301 SIDLKASNGTSLFANENDLKQDLVDLLDVSDFGQSGSETITSQVQKVPASERSLKDR 360
QY 361 VKFKDQOKPRIEKFSIYEYDALSFYSQLELVSKPNSIKDLVNLATLARNLRFSLGKYNF 420
```



```
Db 361 VFPKQDQPRTEKFSLYDYDALSFYSQELVUSKPNISKDVNATLANRFLSLCKNF 420
QY 421 LPDDLASHLDYFLVSKAKIKOSSITKFLFIELPKISLKSSILGDQEPNIKTLFEKVT 480
Db 421 LPDDLASHLDYFLVSKAKIKOSSITKFLFIELPKISLKSSILGDQEPNIKTLFEKVT 480
QY 481 FKLDNPRDVEIEKAFGLLPGVNEEELQARKORASFEKEKSKGLKEFSQOKEENSKAI 540
Db 481 FKLDNPRDVEIEKAFGLLPGVNEEELQARKORASFEKEKSKGLKEFSQOKEENSKAI 540
QY 541 NNQEGLEEDNTERLPENSPLOYQENAGLAGSPDKPMIKDVONORYLAKSOIQELI 600
Db 541 NNQEGLEEDNTERLPENSPLOYQEKAGLSSDPKPMIKDVONORYLAKSOIQELI 600
QY 601 KAKDYTKLAKLSNRHTYINISRLKEQLFDVNPRIPISSRDIIEKAFVLDKTEKNKYWQIY 660
Db 601 KAKDYTKLAKLSNRHTYINISRLKEQLFEVNPRIPISSRDIENAKFVLDKTEKNKYWQIY 660
QY 661 SSASVPFQNKWSLFGYRYRLLGLDPKQTHLVLKQKAGLOPEGYENILPSDFNLEDLKN 720
Db 661 SSASPAFQNKWSLFGYRYRLLGLDPKQTHLVLKQKAGLOPEGYENILPSDFNLEDLKN 720
QY 721 IRIKTPLESQNFKLSLLDPNNYDGEIKAPFGLPLFLPKELRRSNSGSSQNSNSP 780
Db 721 IRIKTPLESQNFKLSLLDPNNYDGEIKAPFGLPLFLPKELRRSNSGSSQNSNSP 780
QY 781 WEQEIISQFQDNLSQDQLAQFSTKIWEKIGDENEFDQNNRLOYKLLKDLQESWINT 840
Db 781 WEQEIISQFQDNLSQDQLAQFSTKIWEKIGDENEFDQNNRLOYKLLKDLQESWINT 840
QY 841 RDNLYWYTLGDKLKVPKPNLEAKFQISNLQELLTAFYTSAAALSNNWYQDSGAKSTI 900
Db 841 RDNLYWYTLGDKLKVPKPNLEAKFQISNLQELLTAFYTSAAALSNNWYQDSGAKSTI 900
QY 901 IFBEIAELDPKVEKKGADYVQLKEHVAIGFDNAGKFNQEVIRSSRTIYLTSGSKSL 960
Db 901 IFBEIAELDPKVEKKGADYVQLKEHVAIGFDNAGKFNQEVIRSSRTIYLTSGSKSL 960
QY 961 EADTIDQNAQVNAPIGLQSFYLDTERFQVFKLATSIAVQHKQKEKTLPKKLNDGYT 1020
Db 961 EADTIDQNAQVNAPIGLQSFYLDTERFQVFKLATSIAVQHKQKEKTLPKKLNDGYT 1020
QY 1021 LIHDKLKKPVIPOISSPEKOWEGKLNQNGSQNVNSTFGSIIIESPYSTNFQEDADL 1080
Db 1021 LIHDKLKKPVIPOISSPEKOWEGKLNQNGSQNVNSTFGSIIIESPYSTNFQEDADL 1080
QY 1081 DQGGQDSRGQNSLNDQEAAGLKQKLAILLGNQFIQYQONDKETIEFEIINVEKYSLS 1140
Db 1081 DQGGQDSRGQNSLNDQEAAGLKQKLAILLGNQFIQYQONDKETIEFEIINVEKYSLS 1140
QY 1141 FRVEFKLAKTLEONGKTIIRVLSDETMSLI VNTTIEKTPMSAVPEVPDTKWVEQYDPRTP 1200
Db 1141 FRVEFKLAKTLEONGKTIIRVLSDETMSLI VNTTIEKAPMSAAPEVFDTKWVEQYDPRTP 1200
QY 1201 LAATKTVLKFQKQIPVDGSGNISDKWLASIPLVIHQOMLRLSPVVKITRELGLKTEQOQ 1260
Db 1201 LAATKTVLKFQKQIPVDASGNISDKWLASIPLVIHQOMLRLSPVVKITRELGLKTEQOQ 1260
QY 1261 QQQQQQQQQQPOKAVRKEBELETYNPKDFNLIPLTKAHLRTLSLNVNNDPNYKIEDL 1320
Db 1261 QQQQQQ-----QKAVRKEBELETYNPKDFNLIPLTKAHLRTLSLNVNNDPNYKIEDL 1315
QY 1321 KVIKNEAGDHLQAFSLRANNIKRLMNTPIITFADYNPFYFNEEDWRSIDKYLNNKNVSS- 1379
Db 1316 KVIKNEAGDHLQAFSLRANNIKRLMNTPIITFADYNPFYFNEEDWNRNDKYLNNKNVSSQ 1375
QY 1380 ----HQQAAGGQSGSLIQRLNKNKIPETFTPALIALKDRNTNLSNYSDKILIMKPKY 1435
Db 1376 QQQQQQQQPPGGGQSGSLIQRLNKNKIPETFTPALIALKDRNTNLSNYSDKILIMKPKY 1435
QY 1436 LVERSIGVPMSTGLDGYIGSEOTKDGTSSSSQQKGFQDQFIQALGLKNTYHGLGLSIR 1495
```

```
Db 1436 LVERSIGVPMSTGLDGYIGSEQLKGGTSSNG-QKRFKODFIQALGLKNTYHGLGLSIR 1494
QY 1496 IFDQNELAKIKDASNKKEGKILKSYDLFQNYLNEYEKSPKIAKGWNTNHPDOKEYPN 1555
Db 1495 IFDQNELAKIKDASNKKEGKILKSYDLFQNYLNEYEKSPKIAKGWNTNHPDOKEYPN 1554
QY 1556 PNOKLPENYLNVLNQPWKVTLVNSSDFITMLFVEPEGSDRGSGTKLKQVIOKQVNNNYA 1615
Db 1555 PNOKLPENYLNVLNQPWKVTLVNSSDFITMLFVEPEGSDRGSGTKLKQVIOKQVNNNYA 1614
QY 1616 DWGSAYLTFWYDKMIITNQPNVITANIADVPIDKVKELNEDNTKLIAPNITQWPNISGSK 1675
Db 1615 DWGSAYLTFWYDKMIITNQPNVITANIADVPIDKVKELNEDNTKLIAPNITQWPNISGSK 1674
QY 1676 EKFKYPTVPFGNWNENSSMNSQOQTPTEWKIREGFALQALKSSFDQKTRTFVLTTNAPL 1735
Db 1675 EKFKYPTVPFGNWNENSSMNSQOQTPTEWKIREGFALQALKSSFDQKTRTFVLTTNAPL 1734
QY 1736 PLWKYGPGLGFQNGPNFKTQDWRLVFNQDNDQIAALRVQEQDRPEKSSDKQKWKIFKV 1795
Db 1735 PLWKYGPGLGFQNGPNFKTQDWRLVFNQDNDQIAALRVQEQDRPEKSSDKQKWKIFKV 1794
QY 1796 VIPEMFNSGNIRFVGVMQIOGPNTLMLPVINSSVIYDFYRGTDGSDNVANLNVAPQVK 1855
Db 1795 VIPEMFNSGNIRFVGVMQIOGPNTLMLPVINSSVIYDFYRGTDGSDNVANLNVAPQVK 1854
QY 1856 TTAFTNNAFNNVKEFNISKKIVE 1879
Db 1855 TTAFTNNAFNNVKEFNISKKIVE 1878

RESULT 3
Q719M1 PRELIMINARY; PRT: 1871 AA.
ID Q719M1;
AC Q719M1;
DT 05-JUL-2004 (TremBLrel. 27, Created)
DT 05-JUL-2004 (TremBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TremBLrel. 27, Last annotation update)
DE P216.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Beaufort;
RA Wilton J.L., Cordwell S.J., Minion F.C., Guerreiro N., Hahn M.,
Walker M.J., Djordjevic S.P.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF541877; AAQ11369.1; -.
SQ SEQUENCE 1871 AA; 214796 MW; CAAAA5PPE1CCE9BB CRC64;

Query Match 96.6%; Score 9402; DB 2; Length 1871;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1819; Conservative 22; Mismatches 28; Indels 12; Gaps 3;

QY 1 MNKKSTLLATAAALIGSTVGVGLASKYKVRGVNPTQGVISQGLIDSVAFKPSIA 60
Db 1 MNKKSTLLATAAALIGSTVGVGLASKYKVRGVNPTQGVISQGLIDSVAFKPSIA 60
QY 61 NFTSDYQSVKALLNGKTFDPKSSSEFTDFVGFDFLTNNGRVLEIPKKYQVVISSEFSPE 120
Db 61 NFTSDYQSVKALLNGKTFDPKSSSEFTDFVGFDFLTNNGRVLEIPKKYQVVISSEFSPE 120
QY 121 DKERFRLQFHLKEKLEDGNIQAQSAKFIYTLPLQMPKALGOYSYIVDKNFNNLIHPL 180
Db 121 DKERFRLQFHLKEKLEDGNIQAQSAKFIYTLPLQMPKALGOYSYIVDKNFNNLIHPL 180
QY 181 SNFSAQSIKPLATLRSSDFIAKLNQFNQDELWVYLEKFEFDLEALKANRLQTADFSFEK 240
Db 181 SNFSAQSIKPLATLRSSDFIAKLNQFNQDELWVYLEKFEFDLEALKANRLQTADFSFEK 240
QY 241 GNLVDPFVYSFIRNPONQKEWASDLNQDKTVRLYLRTFESPOAKTILKDYKYKDTFELS 300
```

Db 241 GNLVDFVYFTRPNQOKEMASDLNQDQKTVRLYLRTFSPQAKTILKDYKXDTEFLS 300
Qy 301 SIDLKASNGTSLFANENDLKQDLVDLDVSDYFGQSETTITSNSQVKPVPASERSLKDR 360
Db 301 SIDLKASNGTSLFANEDDLKQDLVDLDVSDYFGQSETTITSNSQVKPVPASERSLKDR 360
Qy 361 VKFKDQKQKPRIEKFSLEYDALSPYSQLELVSKPNSIKOLVNATLARNRFLSGKYNF 420
Db 361 VKFKDQKQKPRIEKFSLEYDALSPYSQLELVSKPNSIKOLVNATLARNRFLSGKYNF 420
Qy 421 LPDDLASHLDVYFLVSKAKIKOSSITKKLFTLPIKISLKSIILGDOEPNITKTLPEKEVT 480
Db 421 LPDDLASHLDVYFLVSKAKIKOSSITKKLFTLPIKISLKSIILGDOEPNITKTLPEKEVT 480
Qy 481 PKLDNFRDVEIEKAFGLLYPGVNELEQAARASFEKESKIKGLKEFSQKKEENSKAI 540
Db 481 PKLDNFRDVEIEKAFGLLYPGVNELEQAARASFEKESKIKGLKEFSQKKEENSKAI 540
Qy 541 NNQGLEEDDNIATERLPENSPITQYOENAGLGPDPKPYMKDVONQRYYLAKSQIOELI 600
Db 541 NNQGLEEDDNIATERLPENSPITQYOENAGLGPDPKPYMKDVONQRYYLAKSQIOELI 600
Qy 601 KAKDYTKLAKLSNRHTYNISLRKEOLFVNPRIPSSRDIEKAFVLDKTEKNKYQIY 660
Db 601 KAKDYTKLAKLSNRHTYNISLRKEOLFVNPRIPSSRDIEKAFVLDKTEKNKYQIY 660
Qy 661 SSASPVFQKNSLFGYRYLLGLDPKQTHIELVKLGQKAGLQFEGYENLPDFNLEDLKN 720
Db 661 SSASPVFQKNSLFGYRYLLGLDPKQTHIELVKLGQKAGLQFEGYENLPDFNLEDLKN 720
Qy 721 IRIKTPLSQKDNFKLSLLDFNYYDGEIKAPEFGLPLFLPKELRRNSNSGGSONSNP 780
Db 721 IRIKTPLSQKDNFKLSLLDFNYYDGEIKAPEFGLPLFLPKELRRNSNSGGSONSNP 780
Qy 781 WEQEIISQFKQNLQNLQDLQAFSTKIWEKIIGDNEFDQNNRLOYLKDLQESWNKT 840
Db 781 WEQEIISQFKQNLQNLQDLQAFSTKIWEKIIGDNEFDQNNRLOYLKDLQESWNKT 840
Qy 841 RDLNLYTWYLGDKLVKPKNNLEAKPRQISNLQELLTAFTYTSAAASNWNYYQDSGAKSTI 900
Db 841 RDLNLYTWYLGDKLVKPKNNLEAKPRQISNLQELLTAFTYTSAAASNWNYYQDSGAKSTI 900
Qy 901 IPEETAEIDLPKVEKVGADVYQKPHYAI GPDNDNAGKQNEVIRSSRTIYIKTSGSKL 960
Db 901 IPEETAEIDLPKVEKVGADVYQKPHYAI GPDNDNAGKQNEVIRSSRTIYIKTSGSKL 960
Qy 961 EADTTIDQLNQA VKNAPLGLQSPYLDTERFGVFPQKATSLAVQHKQKEXTLPKLLNNDGYT 1020
Db 961 EADTTIDQLNQA VKNAPLGLQSPYLDTERFGVFPQKATSLAVQHKQKEXTLPKLLNNDGYT 1020
Qy 1021 LIHDKLKPVIPOIISSSPEKQWFEGLKQNGQSONVNYSTFGSIIESPYFTNFQEDADL 1080
Db 1021 LIHDKLKPVIPOIISSSPEKQWFEGLKQNGQSONVNYSTFGSIIESPYFTNFQEDADL 1080
Qy 1081 DQGGQDSDRQNGNSLDNQSBAGLLKQKLAILLGNQFIQYQNDKEIEFEIINVEKVSLS 1140
Db 1081 DQGGQDSDRQNGNSLDNQSBAGLLKQKLAILLGNQFIQYQNDKEIEFEIINVEKVSLS 1140
Qy 1141 FRVEFKLAKTLDDNGKTRVLSDETMSLIVNTTIEKTPMSAVPVPDVKWVQYDPRTP 1200
Db 1141 FRVEFKLAKTLDDNGKTRVLSDETMSLIVNTTIEKTPMSAVPVPDVKWVQYDPRTP 1200
Qy 1201 LAAKTKFVKPQKQIPVDSGNSIDKWLASIPLVIHQQMLRSLPVVKTIRELGLKTEQQQ 1260
Db 1201 LAAKTKFVKPQKQIPVDSGNSIDKWLASIPLVIHQQMLRSLPVVKTIRELGLKTEQQQ 1260
Qy 1261 QQQQQQQQQQPOKKA VREKEELTYNPKDEENILNPLTKAHLRSLNVLNNDPNYKIEDL 1320
Db 1261 QQ-----QKKA VREKEELTYNPKDEENILNPLTKAHLRSLNVLNNDPNYKIEDL 1311
Qy 1321 KYIKNEAGDHQLEFSLRANNIKRLMNTPTITFADYNPFYFVFNEDWRSIDKYLNNKGNVSS- 1379
Db 1312 KYIKNEAGDHQLEFSLRANNIKRLMNTPTIXFADYNPFYFVFNEDWRSIDKYLNNKGNVSSQ 1371

Qy 1380 -HQQAAGGQGGSLIQRLNKNIKPETFPTPALIALKDRNNNTNLSYSDKIMIKPKYLVE 1438
Db 1372 QQQQAAGGQGGSLIQRLNKNIKPETFPTPALIALKDRNNNTNLSYSDKIMIKPKYLVE 1431
Qy 1439 RSTGVPHSTGLDGYIGSEOTKDGTSSTSSSQKGFDDPFIQALGLKNTYHKGKLSIRFD 1498
Db 1432 RSTGVPHSTGLDGYIGSEQLKGGTSSNSQV-GFKQDFIQALGLKNTYHKGKLSIRFD 1490
Qy 1499 PGNELAKIKDASNKKGEEKLLKSYDLFPKNYLNEYEKSPKIAKGWNTNIHPDQKEYPNPQ 1558
Db 1491 PGNELAKIKDASNKKGEEKLLKSYDLFPKNYLNEYEKSPKIAKGWNTNIHPDQKEYPNPQ 1550
Qy 1559 KLPENYLNVLNQPKVTLTLYNSSDFITNLVPEPEGSDRGSGTKLKQVIQKQVNNNYADWG 1618
Db 1551 KLPENYLNVLNQPKVTLTLYNSSDFITNLVPEPEGSDRGSGTKLKQVIQKQVNNNYADWG 1610
Qy 1619 SAYLTFWYDKNIITNQPNVITANIADVFIKDVKELEDNTKLIAPNITQWMPNINGSKEKF 1678
Db 1611 SAYLTFWYDKNIITNQPNVITANIADVFIKDVKELEDNTKLIAPNITQWMPNINGSKEKF 1670
Qy 1679 YKPTVFFGNWENENSSMSQAQTPTWEKIREGFALQALKSSFOKTRTFVLTITNAPLPLW 1738
Db 1671 YKPTVFFGNWENENSSMSQAQTPTWEKIREGFALQALKSSFOKTRTFVLTITNAPLPLW 1730
Qy 1739 KYGELGFGQNGFNFKTDQWRLVFNQDNDQIAALRVQEQDRPEKSESDKDKQWIKFKVVI 1798
Db 1731 KYGELGFGQNGFNFKTDQWRLVFNQDNDQIAALRVQEQDRPEKSESDKDKQWIKFKVVI 1790
Qy 1799 EEMPNNSNIIPFVGMOIQGPNTLWLPVINSVLYDFYRGCTGDSNDVANLVAPQVKTIA 1858
Db 1791 EEMPNNSNIIPFVGMOIQGPNTLWLPVINSVLYDFYRGCTGDSNDVANLVAPQVKTIA 1850
Qy 1859 FTNNAFNNVFEFNISKIIVE 1879
Db 1851 FTNNAFNNVFEFNISKIIVE 1871
RESULT 4
Q9KX9 PRELIMINARY; PRT; 560 AA.
ID Q9KX9
AC Q9KX9
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE YX2 (Fragment).
GN Name=YX2;
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISU 232;
RA Liao X., Papazisi L., Geary S.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF279292; AAF87781.1; -.
FT NON TER 560 560
SQ SEQUENCE 560 AA; 63805 MW; C1A89B69D7C04A18 CRC64;
Query Match 28.9%; Score 2816; DB 2; Length 560;
Best Local Similarity 99.8%; Pred. No. 1.2e-107;
Matches 559; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MNKKSTLLATAAAIIGSTVFGTVGLSKVYRGVNPQTQGVISQIGLIDSVAFKPSIA 60
Db 1 MNKKSTLLATAAAIIGSTVFGTVGLSKVYRGVNPQTQGVISQIGLIDSVAFKPSIA 60
Qy 61 NFTSDYQSVKALLNGKTFDPKSEPTDFVSKDFLTNNGRTVLEIPKKYQVWVSESP 120
Db 61 NFTSDYQSVKALLNGKTFDPKSEPTDFVSKDFLTNNGRTVLEIPKKYQVWVSESP 120
Qy 121 DDKERFLRGLKLEKLDGNGIAQSAKFYILLPLDMPKAALGOYSYIVDKNFNLIHPL 180

Db 121 DKERFRLGHLKEKLEDGNIQAQATKFIYLLPLDMPKAAALGOYSYIVDKNFNLIHPL 180

QY 181 SNFSAQSIKPLALTRSSDFIAKLNQFNNDLWVYLEKFFDLEALKANIRLQADSFSEK 240

Db 181 SNFSAQSIKPLALTRSSDFIAKLNQFNNDLWVYLEKFFDLEALKANIRLQADSFSEK 240

QY 241 GNLVDPFVYSFIRNPQKQEWASDLNODQKTVRLYLRTFSPQAKTILKDYKYKDETFLS 300

Db 241 GNLVDPFVYSFIRNPQKQEWASDLNODQKTVRLYLRTFSPQAKTILKDYKYKDETFLS 300

QY 301 SIDLKASNGTSLFANENDLKQDLVDLLDVSDYFGQSQETITSSQVQKVPVPSASERSLKDR 360

Db 301 SIDLKASNGTSLFANENDLKQDLVDLLDVSDYFGQSQETITSSQVQKVPVPSASERSLKDR 360

QY 361 VKFKDQKQKPRIEKFSLEYDALSFYSQQLVSKPNSIKDLVATLARNLRFSLGKYNF 420

Db 361 VKFKDQKQKPRIEKFSLEYDALSFYSQQLVSKPNSIKDLVATLARNLRFSLGKYNF 420

QY 421 LPDDLASHLDYTVLVSQAKIKQSSITTKLFIELPIKISLSSITLGDQEPNITLFEKEVT 480

Db 421 LPDDLASHLDYTVLVSQAKIKQSSITTKLFIELPIKISLSSITLGDQEPNITLFEKEVT 480

QY 481 FKLDNFRDVEIEKAFGLLYPGVNEELEQAKQARASFEKSKKGLKEFSQOKEENSKAI 540

Db 481 FKLDNFRDVEIEKAFGLLYPGVNEELEQAKQARASFEKSKKGLKEFSQOKEENSKAI 540

QY 541 NNOEGLEEDDNIATERLPENS 560

Db 541 NNOEGLEEDDNIATERLPENS 560

RESULT 5

Q9KX7 PRELIMINARY; PRT; 560 AA.

ID Q9KX7

AC Q9KX7

DT 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DE YX2 (Fragment)

GN Name=YX2;

OS Mycoplasma hyopneumoniae.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI_TaxID=2099;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=J;

RA Liao X., Papazisi L., Gearty S.;

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF279293; AAF87783.1; -.

FT NON TIR 560

SQ SEQUENCE 560 AA; 63810 MW; 88AAC16175B7AE97 CRC64;

Query Match 28.5%; Score 2770; DB 2; Length 560;

Best Local Similarity 98.0%; Pred. No. 9e-106;

Matches 549; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 MNKKSTLLATAAAIIGSTVFGTVGLSKVKYRGVNPQTQVISOGLIDSVAFKPSIA 60

Db 1 MNKKSTLLATAAAIIGSTVFGTVGLSKVKYRGVNPQTQVISOGLIDSVAFKPSIA 60

QY 61 NPTSDYQSVKALLNGKTDPKSEFTDFVSKDFLTNNGRVLETPKYYQVVISSESP 120

Db 61 NPTSDYQSVKALLNGKTDPKSEFTDFVSKDFLTNNGRVLETPKYYQVVISSESP 120

QY 121 DDKERFRLGHLKEKLEDGNIQAQATKFIYLLPLDMPKAAALGOYSYIVDKNFNLIHPL 180

Db 121 DDKERFRLGHLKEKLEDGNIQAQATKFIYLLPLDMPKAAALGOYSYIVDKNFNLIHPL 180

QY 181 SNFSAQSIKPLALTRSSDFIAKLNQFNNDLWVYLEKFFDLEALKANIRLQADSFSEK 240

Db 181 SNFSAQSIKPLALTRSSDFIAKLNQFNNDLWVYLEKFFDLEALKANIRLQADSFSEK 240

QY 241 GNLVDPFVYSFIRNPQKQEWASDLNODQKTVRLYLRTFSPQAKTILKDYKYKDETFLS 300

Db 241 GNLVDPFVYSFIRNPQKQEWASDLNODQKTVRLYLRTFSPQAKTILKDYKYKDETFLS 300

QY 301 SIDLKASNGTSLFANENDLKQDLVDLLDVSDYFGQSQETITSSQVQKVPVPSASERSLKDR 360

Db 301 SIDLKASNGTSLFANENDLKQDLVDLLDVSDYFGQSQETITSSQVQKVPVPSASERSLKDR 360

QY 361 VKFKDQKQKPRIEKFSLEYDALSFYSQQLVSKPNSIKDLVATLARNLRFSLGKYNF 420

Db 361 VKFKDQKQKPRIEKFSLEYDALSFYSQQLVSKPNSIKDLVATLARNLRFSLGKYNF 420

QY 421 LPDDLASHLDYTVLVSQAKIKQSSITTKLFIELPIKISLSSITLGDQEPNITLFEKEVT 480

Db 421 LPDDLASHLDYTVLVSQAKIKQSSITTKLFIELPIKISLSSITLGDQEPNITLFEKEVT 480

QY 481 FKLDNFRDVEIEKAFGLLYPGVNEELEQAKQARASFEKSKKGLKEFSQOKEENSKAI 540

Db 481 FKLDNFRDVEIEKAFGLLYPGVNEELEQAKQARASFEKSKKGLKEFSQOKEENSKAI 540

QY 541 NNOEGLEEDDNIATERLPENS 560

Db 541 NNOEGLEEDDNIATERLPENS 560

RESULT 6

Q8KMK1 PRELIMINARY; PRT; 1405 AA.

ID Q8KMK1

AC Q8KMK1

DT 01-OCT-2002 (Tremblrel. 22, Created)

DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Lpps protein.

GN Name=lpps;

OS Mycoplasma conjugativae.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI_TaxID=45361;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=HRC/581;

RX MEDLINE=22464852; PubMed=12576592; DOI=10.1099/mic.0.25864-0;

RA Bellocy L., Vilei E.M., Giacometti M., Frey J.;

RT "Characterization of lpps, an adhesin of Mycoplasma conjugativae."

RL Microbiology 149:185-193(2003).

DR EMBL; AJ318939; CAC87274.1; -.

SQ SEQUENCE 1405 AA; 152862 MW; D9016BB685051F936 CRC64;

Query Match 5.4%; Score 523; DB 2; Length 1405;

Best Local Similarity 22.5%; Pred. No. 3.7e-13;

Matches 298; Conservative 184; Mismatches 488; Indels 356; Gaps 56;

QY 1 MNKKSTLLATAAAIIGSTVFGTVGLSKVKYRGVNPQTQVISOGLIDSVAFKPSIA 60

Db 1 MENKK--LIIGLGAAGVALFAIPVGIASITTKYDG-DPLKNVEEKANSISNIVFDESI 57

QY 61 NPTSDYQSVKALL-NG-KTFDPKSEFTDFVSKDFLTNNGRVLEIPKYY---QVVIS 115

Db 58 GUNTVYASLKSXLVENGHKSDVTAAGFPD-----FTTKNQPTLITSEWNCCEIVIK 112

QY 116 EFPSPDKERFRLGHLKEKLEDGNIQAQATKFIYLLPLDMPKAAALGOYSYIVDKNFNLI 175

Db 113 DIADDDQDQSFVYVYVQSKLDNDVAKSNLKVTKLSFGFLPEFSLASFAHKGQQLFKSI 172

QY 176 ITHPLSNFA-----QSIKPLATRSSDFIAKLNQFNNDLWVYLEKFFDLEALK 227

Db 173 RPRTYGEFTATLKQAENKSVVPLETSSNDNFQAKVKNQAQSDAQQIWNKYPDLOSVID 232

QY 228 NTRLQADSFSEKGNLVDPFVYSFIRNPQKQEWASDLNODQKTVRLYLRTFSPQAKTI 287

Db 233 RIASDTNNVFDNNGVTISKYKSLIKNPNTNNEYALPLVDNKDEVVLFLQTFESDQFKT 292

QY 288 L-----KDYKDETFELSSIDLKASNGTSLFANENDLKQDLVDLLDVSDYFGQSQETIT 342

Db 293 ITNFDGDKFKHIDILRLNLKDLITS-----TKLEDKISITPFSVQDFYSRKDTNP 343

```
QY 343 SNSQVKVPASERSLKDRVK---FKDQOKP---RIEKFSLYEYDALSFYQLOELVSKP 396
DB 344 FTIATKEQNPFANILNEFVKQNIKADFNKPGCTTRVRSPLISKSAI---TQSVEDTSVP 400
QY 397 NSI-----KOLVNT-----LARNRFSLGKYNFLFDDLASHLD 430
DB 401 QTSSAGATGAGATTTPKTTADATLNDPTLAASDISALANQVR--TAKDFDLVDPFIEFIQ 458
QY 431 YVFLVSKAKIKQSSITKK---LFTELPIKISLSSILGDOEPNIKTLFEKEVTFKLDNF 486
DB 459 P-----INGSLFNKEDRDQFOTFVVQOGLNQDLA--KYIRFETGLENLPLGSDVNF 508
QY 487 RDVEI---EKAFGLLYPG-----VNEBLEQARKAQRASFPEKS-----KGLKEPFSQ 531
DB 509 LNPDTLDSQLDLVLYDGRPVKVGTVIVAKSNAGTIIFTKEFSTILGAFKNSLFTISK 568
QY 532 QKEENSKAINNOEGLEEDNITERLPENSPIQ-YQENAGLGASDPKPYMKDVQNRYY 590
DB 569 EKYASL-----PDNI--RLSLNKIRQGLQNNNYTRDASSINEYAYLDLASGK-- 613
QY 591 LAKSOIQELIKAKDYTKLAKLLSNRHTYNISLRKEQLFDVNP--IPSSRDIEKAKFVL 648
DB 614 ----EIDQLIKNSKFDQKLSDRSRYGYQFNNOQILKALTQKFLPSVDF-ATSKR 668
QY 649 DKTERKNYQIYSSASPVFQNKWSLFGYRYLLGLDPRQTTHELVKLGQKAGLOFEGYEN 708
DB 669 SNLEQETNYKILNFNDYFTNDFDVARFY-----TKLAHQGAET 707
QY 709 LPSDF-----NLEDLKNRIKTP--LPSQ-----KONFKLSLLDPNNYDGEIKAPE 753
DB 708 VANYPIELLKANNLID--SSVEISNPKQLFQQLRQISFKDGAQFOYFSFNRYQ---RYSN 763
QY 754 FGL--PLFLPKELRNSN-----SGGSQNSNSPWEQIISQDKONLSNQDQLAQF 803
DB 764 LGFISTLSLQGLRNVAVEILDIAVDGTTQVSPQGTQAVITYSTAKIA-- 815
QY 804 STKIWEKIIGENEPDQNNRLOYLKLLDQLESWINKTRDNLTYWTLGDKLVKPKNNLEA 863
DB 816 SAVDTNEIINFIRYAQNN-----ADPDE-----LSQDL-----NSSA 848
QY 864 KFRQISNLOELLTARY-----TSALSNWNYYQ----- 892
DB 849 GV-EIANLGDLVAFYYKIFSQQFTDLLPISSSLTYNTHFSSQITVDELEAKKAAEAAA 907
QY 893 -----DS-----GAKSTIIPFEEIAELDPKVKKEKVGADVQLKPHYA 928
DB 908 EKIIDAIKRTNTSSTGIDSEFVDIDGSKNALIIQEVL-----KEKDFVKKHVRLLOQNY 963
QY 929 IGF-DDNAGKFNQEVIRSS-SRTIYLYKTS-GSKLEADTIIDQLNQAVKAPLQSFYLD 985
DB 964 LGFKDEKTNVNTFISDYNLLIFVKESDGG--QFNTSQLDQVWIGIPIYKNVYLS 1020
QY 986 TERFGVFOKLATSLAVQHKEKTLPKKLNNDGYTLIHD----- 1024
DB 1021 EEE---YNNKLDTV-----KEIEDKANKTGITDVKDIESVGTSLVDYLKAINPNYN 1069
QY 1025 -KLKKVPVPIQSSSPKDWFEKG---LNQNGQSONVNVSTFG-SITESPYF----- 1070
DB 1070 LVAKRIVLKDGLQPDNGLEGATSLVDSTSPFTNKQKILSLYLTDSPFEDSKSIDNQ 1129
QY 1071 ---STNF-----QEDADLD-----QBGQDSDRQGNNSL 1095
DB 1130 LFLNTYSSIPLIIHKAKEFKLPVNETLAQEQTDQNTQSSPNTQGGTSGNGNGTST 1189
QY 1096 DNOEAG 1101
DB 1190 NQQSAG 1195
```

RESULT 7
Q8G1Z2
.ID Q8G1Z2 PRELIMINARY; PRT; 1427 AA.

```
AC Q8G1Z2;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Lipoprotein.  
GN Name=Ipps;  
OS Mycoplasma conjunctivae.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI_TaxID=45361;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22526333; PubMed=12676664;  
RX DOI=10.1128/JAEM.69.4.1913-1919.2003;  
RA Belloy L., Janovsky M., Vilei E.M., Pilo P., Giacometti M., Frey J.;  
RT "Molecular epidemiology of Mycoplasma conjunctivae in Caprinae:  
Transmission across species in natural outbreaks.";  
RL Appl. Environ. Microbiol. 69:1913-1919(2003).  
DR EMBL; AJ514404; CA555813.1; -.  
KW Lipoprotein.  
SQ SEQUENCE 1427 AA; 155548 MW; 9155221EC1A91B7C CRC64;  
  
Query Match 5.2%; Score 502.5; DB 2; Length 1427;  
Best Local Similarity 20.9%; Pred. No. 2.6e-12;  
Matches 283; Conservative 185; Mismatches 478; Indels 407; Gaps 48;  
  
QY 1 MNKKSTLLATAAAIIGSTVFGTVVGLASKVKYRGVNPQTQGVISQGLDLSVAFKPSIA 60  
DB 1 MENKK--LIIGLGAAGVALFAIPVGIASTTKYDG-DPLKNVEEKVNSISNIVKDESI 57  
QY 61 NFTSDYQSVKALL-NG-KTFDPKSEPTDFVSKFDELTTNGRTVLEIPKY---QVVIS 115  
DB 58 GVNTTYASLSKLVENGHKSDVTAFFGDFD-----FTTKNQPTLITDSEWNCCEIVIK 112  
QY 116 EFPEDDKERFRGLGKLEKEDNIAQASATKFTYLLPLDMPKAAALQYGYVVDKNFNNL 175  
DB 113 DIADDDQDQFKIYYVQSKLDNNDIAKSNLRVKTLSFGFLPFESLASPAHKGQOLFKSI 172  
QY 176 IIHPLSNFSA-----QSIKPLATRRSDFIAKLNQFNNDQELWYLYLEKFFDLEAKA 227  
DB 173 RPTYGFIATLKGAEKNSVPLELTSSNDFAKDNVKAQSSDEAQQIINKYFDLQSVID 232  
QY 228 NIRLOTADPFEGKLVDPFVYFIRNPQKQEWASDLNODOKTVRLYLYLTFESPOAKTI 287  
DB 233 RIASDANNVFDNNGVTISKYKSLIKNPNTNBYVPLVDNKKDEWVLFLOTEFSDQPKET 292  
QY 288 L-----KDYKYKDBETFLSSIDLKASNGTSLFANENDLKDQDLVDLDDVSYFGQSEIT 342  
DB 293 ITNFDGKDFKHIDILRLNLKDLTS-----TKLEDKISITPFSVQDFYRRDKTNP 343  
QY 343 SNSQVKVPASERSLKDRVK---FKDQOKPRIEKFSLYEYDALSFYQLOELVSKNSI 399  
DB 344 FTIATKEQNPFANILNEFVKQNIKADFNKPGCTTRVRSPLISKSAI---TQSVEDTSVP 403  
QY 400 KD-----LVNATLARNLRFSL-----GKYNF----- 420  
DB 404 QNTGAGAGTTTTTTTADATLSPTLAASDISALANQVRTAKDFRLVDVDFIEPTQFINGS 463  
QY 421 LFDLASHLDYFVLVSKAKIKQSSITKKLFI-----LPI-----KISLKS----- 462  
DB 464 LFTNKEDRDQFLFVVQOGLNQ-DLTKYIRFETGLENLPLGSDVDFKFLNFDITLDSQLD 522  
QY 463 ILGQEPENIK---TLFKEVTFKLDNPRDVEIEKAFGLLYPGVNEELEQARKAQRASFEK 519  
DB 523 VLYDGRPVWVGTVIVAK-----SNAGTIIFTKEFSTILGAFKNSLFTISKIYASL-- 574  
QY 520 ESKKGLKEFSQKKEENSKAINNQEGLEEDNITERLPENSPIQ-YQENAGLGASDPK 578  
DB 575 -----PDNI--RLSLNKIRQGLQNNNYTRDASSINE 603  
QY 579 YMIKDVQNRYYLAKSQIQELIKAKDYTKLAKLLSNRHTYNISLRKEQLFDVNP--IP 636  
DB 604 YAVLDLASGK-----EIDQLIKNSKFDQKLSDRSRYGYQFNNOQILKALTQKFLVP 657
```

637 SSRDIEKAFVLDKTEKNKYWQIYSSASPVFONKSLFGYRYLLGLDPKQTIHELVLKLG 696
658 SVEDEF-ATKSRSSLEQETNYKILFNNDYFTNDFVARFY----- 697
697 QKAGLOFEGVENLPSPDF-----NLEDLKNIRIKTP--LFSQ-----KDNFKLSLLDF 741
698 --TKLAHQGAETVANYFIELLKANNLID--SSVEISNPKOLFQOLRQISFKDNAKQYFSP 754
742 NNYDGEIKAPBFL--PLFLPKELRNS-----SNSGQSNSNSPW 781
755 NNNYQ---RYSNLGFTSLSLPEGLKAVKILKFAIVDSSTQVSPQGTQPSIPTY 811
782 EQEIIISQFQDNLSNOQLAQFSTKIWEKIIGDNEFDQNNRLOYKLLKDLQESWINKTR 841
812 PTAKIASAVDTN-----EINFIIDYQANN----- 836
842 DNLYWYTLGDKLVKPKNNLEAKFR-----QISNLOELLTAFY-----TSA 882
837 -----VDPDAELSDQLNSSADVEIANJGDLVAFYKIFFSQOFTDTLLPIS 883
883 ALSNNMNY-----YQDSGAK-----STIIPBEIAELDP----- 910
884 SLTYNTHFSQITVDELEAEKAAEAAKILQDAIKKAMTNSGIAAQLADVDDSQNDFID 943
911 ---KVKEKGVADYQVKFHAIGFDDNAGKFNQEVIRSSRTIYLYKTSKSKLEADTIDQ 967
944 GFLSEKDFVKHTRLLQYNSLGFDEKNTENNIFISDNYNLFIFVKENNGKQFNTSGQ 1003
968 LNAQVKNAPLGLSFLYDTERFGVFOKLATSLAVQHKQEKTLPKKLNNDGYTLIHDCLK 1027
1004 LDQVIGTIPTIYKNVLABEE---YNKLDFTV---KEIEEKANTGITDVKD--- 1049
1028 KPVIQISSPEKDWF-----EGKLNQNGSQNVNVTFSIIBSP-- 1068
1050 ---IESIVGTSVLDYLKAINPNVNLVAKRIVLKDGL-QTGDSGLEGASSSILVDSGP 1105
1069 -----YFSTWF-----Q 1075
1106 FANQONKILSLYLTDPNPKDKSIDNQLFLDTNYSSIPLKIIHKAERFKLPVNEVLAQ 1165
1076 EDADLD-----QDQDDSRQGNNSLNDQEAG 1101
1166 EQTDQSTQNSPNTQGGTQSGNGNNGNSTNGOSSG 1198

RESULT 8

Q9KGX3 PRELIMINARY; PRT; 1296 AA.
AC Q9KGX3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Adhesin like-protein p146.
GN Name=p146;
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BQ14;
RA Verdin E., Saillard C.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF279908; AAF91425.1;
SQ SEQUENCE 1296 AA; 146344 MW; 024ECB30C12D9670 CRC64;

Query Match 4.9%; Score 473.5; DB 2; Length 1296;
Best local similarity 21.6%; Pred. No. 3.6e-11;
Matches 309; Conservative 195; Mismatches 486; Indels 443; Gaps 65;
2 KNKSTLLATAAIIIGSTVGVGVGLASKVKVGVNPTQGVISOLGLIDSVAFKPSIAN 61
3 KNKNSLLV-TATAIVGVAVFATTVGLVTRIRYKGENPRAELESLSVKQNVAFKSDVFD 61

62 FTSDYQSVKALLNGKTFDPKS-----SEFTF-----VSKFDLTLNNGRTVLEIPKY 110
62 NSTTYKQI-KALLFDETKLRPGIDLNFIFSYTAVNSKIQKF-----VTFAPNKP 112
111 QVWISFSPEDKFRGLGFLKEKLEDGNIQAQATKFIYLLPLDMPKAALGQVSYIVDK 170
113 PFEFINLPDDKNQFTTLQFRAKHQLDNNYATYSS-----ILSKIAVAQRSQFA-LADF 166
171 NFN-----NL-----IIHPLSNFSAQSIKPLALTRSSDFIAKLQNFQODE 211
167 NANHRKITKSFQTNIQNLRETDFSYDFSSTSLASQKIPFLTRVEDFAADINKSGNOEE 226
212 LWVYLEKFP-----DLEALKAN----- 228
227 AISRISKYFPDFQRYIHELKDDPNVLPFKGKIFDFISITRACGTNDFISLSANSEPSFL 286
229 --IRIQADAFSEKGNL-----VDPFVYSFIRN-----PONQEWASDL 265
287 IKARL-TNEAKFELRGLNIEEAEMLEETKLVPVDQFVVNLETDLKPGQAPEKSQKPSQ 345
266 NODQKTVRLYLRTEPS--POAKTILKDYKYKDETFLSS-IDLKASNGTSLFAN-----EN 317
346 TETKTYFAEIDKILSKITMRKQLQSLDPKAPQTSQSPQVKAS--VSAMSNLDAQOEN 403
318 DLKDOLDVDLDDVDFGQSETITNSQVFPVPAERSLSLQDVKF-----KKDQKQPR 371
404 KILVFNQSSNPQOQOQFQSQPOQFQSQPOQFQSQPOQFQSQPOQFQSQPOQFQSQ 463
372 I----EKFSLEYDALSFYSQLQELVSPKNGIK-----DLVNAT-LARNLRFSLG--KY 418
464 TDPDSKFKIRTKRARDFLKBNKTFYRSNKLKSKQLEKINSEYLSNKGIDGLGVKKY 523
419 NPLFDLASH--LDYFVLSKAKIK--QSSIFKFLFIELPIKISLSSILGQ----- 467
524 -----ISNNOGIEYTFDIANAKIRDAQDGITS--HIETPVTISLWSSPFGSDNVLLKS 575
468 --EPNKTILFEKVTFKLDNPR-----DVEIEKAFGLLY-----PGVN-----BE 505
576 KTETIIPYFQKETTSEKDKQVGHQTKELDNLQKLVQLSELPGTISTOGSSGSTQTQ 635
506 LEQAKARQASR-----EKEKSKGLKEFSQKKEEN-----SKAINNOQGLEEDDNITERLP 557
636 IKEVKPLTITAFISKELEALIDGKNLASQTSQAVSVSQEVKTTEFQOQEANSTNSP 695
558 EN-SPIYOQENAGLQSPDKPMKD---VONQRYLAKSQIOELIKAKDYTKLAKLS 613
696 TSPSP-----SPSPTSPKLDENIGVFNPR-----FEEIKKIIS 729
614 NRHTYNISLRLKEOLFV-----NPRIPSSRDIEKAKFVLDKTEKNKYWQIYSSASPVFONK 670
730 SBTYKYNFRANEALLDAMWGKONFPLKDI--SQRSQRLAKDYKLVNLKSNKFLKED 787
671 WSLFGYRYLLGLDPKQTIHELVLKQKAGLOFEGYENLPSPFNLEDLKN-----RI 723
788 YDVLAFYANLVQKDPREVLYLFEIAKA-----NNLIGPEEKLDLQIEDDGIFRRA 839
724 KTFPLFSQDNFKLSL--LDFNNYDGEIKAPBFG--LPLFLPKELR-----RNSNSGG 773
840 KAIKLDKSNNGGIYGFSENNQF---LKFHERGMSTLYLNEAKTKLADYONLLSQGI 896
774 SQ-----NSNSPWEQEIISQFQONLSDQLAQFSTKIWEKIIGDNEFDQNNRLOY 826
897 SDTKIFSELNKIQLDNLNKAQSSSDSKSDSSDS-----SDAKTPTNQDILLS 946
827 KL--LKDLOESWINKTRDNLWT-----YLG-----DKLVKPKNNLEAKF-- 865
947 KLTSLKSOEIAIVKK-----YETESKKYLGTENNNGSSSGTEQKGSIIPEEN--KKFIL 999
866 ---RQISNIOELLTAFYTSAAASN--NNNYQDSGAKSTIIFEEIA----- 906
1000 ENTAKLDNLADLLAFYQAKRLNFTSWSQLQDDEDLDYQIQFEKEANNTSSSSSSSS 1059
907 -----ELDPKVEKVGADVYQVKFHYAIGFDDNAGKFNQEVIRSSRTI---YLKTSG 956

```
Db 1060 SSSSSSTDTNKP- -NAVEYKLYYYKI - - - - - YNKTTKVYVTPKTIKLYLASSN 1112
QY 957 KSKLEADITDOLNOAVKAPLGLQSFYLDTERFGVFOKLATSLAVQHKQKKTLPKLANN 1016
Db 1113 IGVKEKQRELMNKLVLSPAYSIFYL - - - - - KQSE - - - - - 1144
QY 1017 DGYTLIHDKLKPVIPQISSPPEKDWFEKLNQNGSQSNVNVSTFGSIIESPYFSTNPOE 1076
Db 1145 - - - - - WQVKITNNGQ - - - - - 1155
QY 1077 DADLDQGDSDRQGNNSLDNQAGLLKQKLAILLGNQFIQYQONDKIEFEFINV--E 1134
Db 1156 - - - - - QMGQTSSQGFSELPFK - - - - - KIQEIYVHRNNKDYLVKVVTRDD 1196
QY 1135 KVSELSFVPEKPLAKTLDNG--KTIRVLSDETMSLIVNTTIETKTPMSAVPE 1185
Db 1197 AYAENAKIVHLVRVKEEQGRKTEREVEKEQ - - - - - KETSSQVFPQ 1241

RESULT 9
Q6RS18
ID Q6RS18 PRELIMINARY; PRT; 1078 AA.
AC Q6RS18;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Adhesin.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_taxID=2099;
RN [1]
RP SEQUENCE FROM N.A.
RA Boivert A., Cheikh Saad Bouh K., Khevar A., Shareck F., Dea S.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY512903; AAS00638.1; -.
DR InterPro; IPR000719; Prot_kinase.
DR PROSITE; PS00107; PROTEIN KINASE ATP; UNKNOWN 1.
SQ SEQUENCE 1078 AA; 122102 MW; 717A6B6AB6D4F45F CRC64;

Query Match 4.2%; Score 410.5; DB 2; Length 1078;
Best Local Similarity 20.6%; Pred. No. 1.1e-08;
Matches 259; Conservative 211; Mismatches 48; Indels 313; Gaps 59;

QY 1 MNKKSTLLATAAIIIGSTVFGTVGLASVKYKRGVNPQTQGVISQLGLDSVAFKPSIA 60
Db 1 MSKSKTKPIGLTAGIVGLVGLTVGLSSLAKEYSESPRKIANDFAAKVSTLAFSPYAF 60
QY 61 NFTSDYQVKKALLNGKTFDPKSEFTDFVSKFDFLTNNGRTVLEI - - - - - PKKYQV 112
Db 61 ETDSYKIVKRWLVDSNN-NIRNKE--KVIDSFSPFTKNGQLEKINFPDPEYTKAKITF 117
QY 113 VISEPSPDDKRFRLGPHLKEKLEDGNIAGSATKFIYLLPLDMPKAA--LGQVSYIVD 169
Db 118 EILEIIPDVNQNFVKVQALQKLNHGDIAKSD---IYEQTVAFAQSNLLVAEFLSK 174
QY 170 KNFNML-----IIHPLSNFS---AQSIKPLALTRSDFIKLNQFNQNNQDELWYLEKFF 220
Db 175 KITEKLNQNIENLSKPTINFADKETSQKDPSTLRAIDFYDNLNTARNPEDLDIKLANYP 234
QY 221 DLEALKANI-RLQTADFSEKGNLVDPPVYFIRNPQKQEWASDLNQDQKTVLYLRT 279
Db 235 PV--LKNLINLNAPENKLPNNLGNIFKFSFAKD--SSTNQYVSTQNG---IPSLFLKAD 288
QY 280 FSPQAKTILKDYKYDET--FLSSIDLKASNGTSLFANENLDKQDLVDLLDVSDYFGGQ 337
Db 289 LSQSAREILAS---PDEVQPVINIURLMKKONSSYFLNFDFVNNL - - - - - 331
QY 338 SETITNSQVFPVPAASERSLKDRVKFKKQDQKPRIKFSLEYE - - - - - DALSFY 386
Db 332 - - - - - TLKN---MQKEDLNAGQNLSAVEFLADIKSGFFPGDKRSSH 370
QY 387 --SQQLVSKPNSIKDLVNATLARNLRFSLGKYNFLFDDL--ASHLDYYFLVSKAKIKQ 442
```

```
Db 371 TKSEISNLLNKENIYD - - - - - FGKYNGFNDRNLNSPNLEYSLDAASASLDK 417
QY 443 SSITKKLFIELPKISKLSKSSILGDQ--EPNIK-TLFEKEVTFKLDNFRDVEIEKAFGLLYP 500
Db 418 KD---KSVIUIPYRLETKDKFFADDLYPKKDNILVKEGILKLTGFK---KQPKIDLP 469
QY 501 GVNEELQARKAORASPEKESKKGLKEFSSQKEENSKAINNOGLEBEDDNITERLPENS 560
Db 470 NINQOI - - - - - PKTEY---LPFFEKKEEQAK - - - - - 493
QY 561 PIQYQONNAGLGASPPKPYMIKDVONQY--YLAKSOIQELIKAKOYTKLAKLSNRHTY 618
Db 494 -LDY- - - - - GNILNPSYNTOLAKVEVEALFKGNKQEIYQALDGNAY 535
QY 619 NISLRLEQLFDVNPRI--PSSRDIEK- - - - - AKFVLDKTEKNKYWQIYSS--A 663
Db 536 EFG-AFKSVLNSWTGKIQHPEKADIQRTHLFOVKLGNSVLNQPOTTKE-QVSSLSK 593
QY 664 SPVFQNKWSLFGYRYLLGLDPKQTIHELKVGKAGLOFEGYENLPSDFNLEDLKNI-- 721
Db 594 NNFFKNGHQVASYFQDILLTKDLTVLSTLYDLAKKWGLE-TNRAQFPKEV-FQDTKDIFA 651
QY 722 ---RIKTPFSQKDNF---KLSLLDFNNYDGE-IPAFEGCLPLPLPKELRRNSSSGG 773
Db 652 EADKLFLEWKEKDPYNQIKEIHQLSFNILARNDVIKSDGFGYVLLLPQSVKTELEGKN- 710
QY 774 SQNSNSPWEQEIISQFQDONLSNQDQLAQFSTKIWEKIIIGDENEFQD--NNRLOVKKLKD 832
Db 711 - - - - - EAQIPEALKKYSLIEN---SAFKTILDKNLLEGTDFTKFGDKAFLKAA 759
QY 833 Q---ESWINKTRDNLWYTY---LGDKLKVKPKNNLEAKFRQISN-LQELLTAFYTSAA 883
Db 760 QNNFAPWA-KLDDNLQYSFEAIKKGTTKEGKREEDVKVKELDNKIKGILP--QPPAA 816
QY 884 LSNWNYQSDGAKSTIIPEIAB--LDPKVEKVGCA - - - - - 918
Db 817 - - - - - KPEAAKPVAAKPEAAKPVAAKPEAAKPEAAKPEAAKPEAAKPEAAKPEAAK 868
QY 919 ---DVYOLKPHVAIGFDDNAGKFNQBVIRSSSTIYKLTSGKSKLEAD-TIDOLNOAVK 973
Db 869 KPREDYFPMAFSKLEYTDE---NKLSTKTPINVLVLELVHQSEYEQKIKELDKTVL 924
QY 974 NAPLGLQSFYLDTERFGVFOKLATSLAVQHKQKKTLPKLNNDGYTLIHDKLKPVIPQ 1033
Db 925 NLQYQFQEVKVASDQ---YQKLSPMMTEGSSNQCKKAEGAPNOG-----KKAEGAPS 974
QY 1034 ISSPEKDWFEKLNQNGSQSNVNVST- - - - - FGSIIIESPYFSTNFOEDADLD 1081
Db 975 QGKKAEGAPSGQK-KAEGTSNQQTITTELNYLPDLGKKIDEIHKQ--GKNWKTEVELI 1031
QY 1082 QDG - - - - - QDSSRQGNNSLDNQAGLLKQKLAILLGNQFIQYQONDKIE 1127
Db 1032 EDNIAGDAKLLYFILRDDSKSG-----DPKSSSLKVIITVKQSN-----NNQELK 1076

RESULT 10
Q6RS18
ID Q6RS18 PRELIMINARY; PRT; 1037 AA.
AC Q6RS18;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Adhesin.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_taxID=2099;
RN [1]
RP SEQUENCE FROM N.A.
RA Boivert A., Cheikh Saad Bouh K., Khevar A., Shareck F., Dea S.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY512894; AAS00629.1; -.
DR InterPro; IPR000719; Prot_kinase.
```

DB 816 -----KPEAAKPVAAPKAAPVATNTNTGFSLTNPKPE-----DYFPMAFSTKLEY 864
 QY 932 DNAGKFNQEVIRSSRTIYLKTSKGSKLEAD-TIDOLNOAVKNAPLGQSFLDTERFG 390
 DB 865 TDE----NKLSLTPTEINVLFLVHQSEVEHKKIKELDKTVLNLYQYQFEVKVTSEQY- 919
 QY 991 VFQKLATSIAVQHKKKEKTLPKKLANNDGYTLJHDLKPKPVIPOISSSPKDWFEGLNQ- 1049
 DB 920 |||||-----QKLSHPMTTEGSSNOQKK-AEGAPNOG 945
 QY 1050 ----NCOSQNANVSTF-----GSIIESPFTNFQEDADLDQDG----- 1084
 DB 946 KXAEGASNOQSTTETLNTPLPDGKKIDBIIRKQ--GKWKTEVELIEDNIAGDAKLLYF 1003
 QY 1085 ---QDSRQRGNNSLDNGEALLKOKLAILLGNOFIOYYOONDEIE 1127
 DB 1004 ILRDDSKSG-----DPKSSSLKVKITVKQSN-----NNQELK 1035

RESULT 11
 Q81525 PRELIMINARY; PRT; 5767 AA.
 ID Q81525
 AC Q81525
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN ORFNames=PFL1930w;
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OC NCBI_Taxid=36329;
 RN [1]
 RP MEDLINE FROM N.A.
 RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyse S.,
 RA Chan M.S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
 RA Pertea M., Allen J., Selengut J.J., Haft D., Mather G., Vaidya A.B.,
 RA Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
 RA Fraser C.M., Barrell B.;
 RT "Genome sequence of the human malaria parasite Plasmodium
 falciparum.";
 RL Nature 419:498-511(2002) .
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Hyman R.W., Fung E., Conway A., Kurdi O., Mao J., Miranda M.,
 RA Nakao B., Rowley D., Tamaki T., Wang F., Davis R.W.;
 RE Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AE014850; AAN36472.1; -
 KW HSSP; P40136; iPKO.
 KW Hypothetical protein.
 SQ SEQUENCE 5767 AA; 691409 MW; E84A340441C67160 CRC64;

Query Match 4.2%; Score 405; DB 2; Length 5767;
 Best Local Similarity 19.3%; Pred. No. 1.6e-07;
 Matches 406; Conservative 328; Mismatches 687; Indels 680; Gaps 104;

QY 65 DYOSVKK-----ALINGKTFFPKSSEF-----TFVSKFDFLTNNGRTLVBIPKYQVVIS 115
 DB 3290 EYDSLKKFKDKQKNNEYIDILSQFHKKDIYEKGEKNNSNELSNINK-DVILP 3348
 QY 116 EFSPDDKERFRGLGHUK--EKLEDGNI-AQSATKFYLPLDMPKAAIGOYSIVDK-- 170
 DB 3349 SLISFNRCDIMKLSHDVNSIELNDDLNNKELNVVISLTKTDI-ENMLDNHTNYHNKDND 3407
 QY 171 ----NFWNLIIHLPSNFSQAQSIKPLATRSDSFIAKLNQ-----FNNQDELWLVYL-E 217
 DB 3408 HNNTYNYFNHG-SGFIPYEDNNISKINKODDLIEKLKQKDSIKEYNKLDNNNIYVNK 3466
 QY 218 KFFOLEALKANIRLOTADFGEKGNLVDPFYVSFIKPNQKQEWASDLNQDKTVLYLR 277

Db 3467 KERIEKLGK--EYGDQFSINKN-----SYISNNNDNRNNSNLSEK--HLHIK 3513
Qy 278 TEPSPQAKTLKDYKYKDETFLSSIDLKASNGTSLPANENDLKQDLVDLLDVSDYFGGQ 337
Db 3514 EBY-----KELNVEYKF-----SUNLU-----NTRKNKLD----- 3538
Qy 338 SETITNSQVPVPASERSLDRVKFRKQDOQPRIEFPSLYDAL--SPYSQJQELVSK 395
Db 3539 -----TLK--REKDKMEK--LKELOIHDKDYIIQNLKGINRELENN 3575
Qy 396 PMSIKOLVATLARNLRSFGKYNPL--FDDLASHLOYFLVSKA-KIKQSSYFKLPIE 452
Db 3576 LNEIKSFPNEKI-----LEKENIIEKNLDLNNLBEIIMLSIKIKNNNFIKEL--E 3626
Qy 453 LPIK-----ISLKSILGDQPNLK-----TLFEKEVTFKLONFRDVEIEKAFGLLYPCVN 503
Db 3627 LOIKNLNNEINTLMDLKDSSEIRMLNHTLEBKQ-----N 3662
Qy 504 BELEQARKAQRASP--EKEKSIGKGLKEFSQKKEENSKAINNQBLEBDDNITERLPENS 560
Db 3663 ECVKLRNKLQNYLHISNEDKYDKSSSTYNNNNNNNNNNNNNNNNNSDDEINE----- 3716
Qy 561 PIQVOEQEAGLQSPDKPYMTKDVONQRYYLAKSOIQELIKAKDYTKLAKLLSNR-HTYN 619
Db 3717 -TRRRQDN-----QIDDLKR-----KSVLQSNHEYKQNNELRRDLNNSFSSYN 3758
Qy 620 ISLRLEQOLFDPNPRIPG-SRDIEKAKFVLDTKKNYQWQYSSASPVFQNKWSLFGYR 678
Db 3759 ---KJK---FOSEKQIEDFEKANMLKFYINKIE-----ENK----- 3790
Qy 679 YLLGLDPRQTHIELVKLQCK-----AGLQFGEYENLPSPDF--NL-----ED 717
Db 3791 -----KKEEXHALIEVAIKEMDKNEIKTLQLEKNVFINDIQNLEKKKTLLEYLEED 3844
Qy 718 LKNRIKTPLSQKDNFKLLDFNNYD-----GEIKAPEGLPL-----FLPK 763
Db 3845 IKIVQHIIMKYVOKLN---SLKDIEQLDELNENELKTNEYFTFIKQIFISIHQNYMKEE 3901
Qy 764 LRRSSNGSGGQNSNSPWEQIISQKQNLQSNODLAQFQKIWEKIIGDENFDF--Q 820
Db 3902 LNRK-----EQMLN---CQGLIKENEVIQLEKELKELTL-DKNYRDKDIK 3943
Qy 821 NNRLOYKLLDQLESWINKTRDNLNLYTLGDKLVKPK--NNLEAKFRQISNLQELTAF 878
Db 3944 NQKDLRLHIDRKDSY---EDEL--NHIINELEIKREEINQMNNKLNELKNIQ----- 3991
Qy 879 YTSALSNNNYQDSGAKSTIIPFEEIAELDPKVKYKGVADYQVLKPHYAIGFDNACKP 938
Db 3992 -----NENKRVEDIKT-----EGDLKYKE-----ELKYR-----EENLKH 4025
Qy 939 NOEV-----IRSSRTIYLTGSKSLKLEADTIDQLNO-AVKNAPLGLQSPYL 984
Db 4026 EELLIKENYMNKELMKSEELIVRENDLIKNKEDLIDKTNELMLKRELQKDKIHI 4085
Qy 985 DTERFVGFKLATSIAVQH--KQEKTLPKKLNDNDGYTLIHDKLKPKVIPQISSPEKOW 1042
Db 4086 DELKNDMLKPEKLYNAQNDIKVETKLLK-----IVEKLKDHNIIEI----- 4128
Qy 1043 FEGKLNQNGSQSNVNVSTFGSIIESPYFTNQFQEDADLDQDQDSDRQGNNSLNOEAGL 1102
Db 4129 -----FDINNRDNNNSIKN-----HTSNINSDI-----KNHTTNNNNNNNNNNR 4173
Qy 1103 LKQKLAILLGNQIOYQONDKIEBF---EINVEKVSLSFRVEF-KLAKTLEDNGKT 1157
Db 4174 SLKRLGILLGLDEIG--DENELDPHFQKENSLSLDSLEYTYTFNEFKRIKLNKNN--- 4228
Qy 1158 IRVLSDETMSLIVNTTIETKTPMSAVPFTQKWEQYDPRTPAAKTKFVLKFKDQIPV 1217
Db 4229 --LLNRE-----KQVEIYKKSVEIEKKEIYKIE----- 4256
Qy 1218 DSGNISDKWLASIPLVTH-CQMLRL-----SPVKTIRELGLKTEQOQOQ 1263

Db 4257 -----LEKUNELLHVEQNNRKNLDLELEKYKSEDTTHIVKSLRE-----SEELNE 4301
Qy 1264 QOQOQOQKAVRKEBELETYNPKDEFNLTNPLTKAHLRTLSNL--VNNDPNYKI----- 1317
Db 4302 KNNKILELOQLKLETSYBEINIMIDKKN-NLJKEKDQYQEKIDNLSIINDYEKEIKEMN 4360
Qy 1318 -----EDLKVIKNEAGHQ-LAFSLRANNIKRLMNTTITFADYNPFF 1358
Db 4361 KEKLLTRKSIEMIDGEDIKKLDLDBAQLILDLEKKN-BELNNINIELIN----- 4414
Qy 1359 YYNEDWR-SIDKYLNNKGNVSSHQOQAAAGNQGSLIORLNKIKPETFPTPALIALKORN 1417
Db 4415 -SNKMDRLDIDLVSNDNLNK-----ENILNEKKMNEDEKYNELKLTAYQKV 4462
Qy 1418 NTLNSYSDKIIMIKPKYLVERSIGVPWSTGLDGYIGSEQTKDGTSSSSQOQGFDPFIQ 1477
Db 4463 HEFNKLYNMLSIKMNKFDSENDMEQNYT--NDNVNMQKSRNSINNKYDINDPENDTII 4519
Qy 1478 ALGLKNTYHGKLGSLIRIFDPGNELAKIKDASNKGBEKLKSYDLFKNYLNE----- 1531
Db 4520 ALNEKN-----MIIESLNEKTAHYEENYKKEELVKN---YKMINELSNKIH 4564
Qy 1532 -----YEKKSPIK-----AKGWT 1544
Db 4565 IPEDNLDLISLNEFIKTNFTYQNLIDILNLSILYKLLSKLILVENQFYKIISIRDWL 4624
Qy 1545 NIHPDQKEYPNQKLPENYLNVLNQP-----WKVTLYNSSDFITNLF 1588
Db 4625 FEIIKSISYNSNN--NNNNIISNNTRESVLRMSLRKKGFKRTASVTSIYNLDF 4681
Qy 1589 V-----EPBGSD-RSGTGLKQVIOKQV-----NNY---ADWGSAYLTFWYDK 1628
Db 4682 LSENVLSAEKENDEIMNSFNKTLNVIDDDFNSPENICIELNLFIEEDMNNIFIKL----- 4737
Qy 1629 NIITNOPVITANITADYFIKDVK-----ELEDNTKLIAPNITQWMPNIGS 1674
Db 4738 NILNQ-----FKNIIILIHMKKINKKNNNEKLEKSVKILKRLN-----SISND 4789
Qy 1675 KKPYPKTPVFGWNEENSSMNSQO-----TPWEKIRGFALQAL---KSS 1719
Db 4790 LEKRTBEI-----EHLNIALNKECAEKILLNQYEDIKKTQKQNEQYNEQIILIOKNE 4843
Qy 1720 FDKTRFTVLITNAPLWKYGLPGFONGPNFKTDWRLVFQNDNQI-----AALRVQ 1773
Db 4844 YELKSLSDQLLLE-----KNENOIKTEKINKLNSDLNHNFTSLDLSAYQLS 4890
Qy 1774 EQDRPEKSS-----EDKDK-----QKWIKFKVVIPEEM--FNSGNIRFVGMQIOGNP 1819
Db 4891 ELNNEKKKKNKILLNELEBEKEIISHLQKDPESNVEVINELKEFNEMLINKLKKEEYKKE- 4949
Qy 1820 TMLPVTINSSVIYD-----FYRGTGSDSNVANLNVAP---WQVKTIATFTNNAFNVFKEF 1871
Db 4950 -----KDNVYDKNSHDNYSNNNDNNNNNSNNSSNRSHSNNNNNNNNNNNMKSF 5002
Qy 1872 N 1872
Db 5003 H 5003
RESULT 12
Q6RS10 ID Q6RS10 PRELIMINARY; PRT; 1063 AA.
AC Q6RS10;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Adhesin.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
RN NCBI_Taxid=2099;
RX [1]
RP SEQUENCE FROM N.A.
RA Boilevert A., Cheikh Saad Bouh K., Kheyar A., Shareck F., Dea S.;

[illegible]

```
QY 387 --SQQLSVSPNSIKDLVATLARNLRFSLGKYNFLPDDL--ASHLDYFLVSKAKIKQ 442
DB 371 TKAEISNLNKKENIYD-----FGKYGKENDRLNPNLEYSLDASASLDK 417
QY 443 SSITKKLFIEPIKLSKSSILGDO-EPNIK-TLFEKEVTFKLDNFRDVEIEKAFGLLYP 500
DB 418 KD---KSVLPYRLEIKDFADLDYDKDNILVKEGILKJGFK-----KGPKIDLP 469
QY 501 GWNELEQARQASPEKSKGKLEFSQOKESKAINNQGLEBDDNITERLPNS 560
DB 470 NINQOI-----FKTEY-----LPFEKGEQAKL-----DYGILN--PNNT 505
QY 561 PIYQOENAGLGASDPKPYMIKDVQNRQYLLAKSQIQELIKAKDYTKLAKLSNRHTYNI 620
DB 506 -----QLAKVEALFKGNQKEIYQALDGNAYAYEF 536
QY 621 SLRLKEQLFDVNPRI--PSSRDIEK-----AKFVLDKTEKNKYWQIYS--ASP 665
DB 537 G-AFKSVLNSWTGKIQHPKADIQRTRHLEQVKIGSNSVLNQPTTKE-QVISLSKSN 594
QY 666 VFONKWSLFGYRYLLGLDPKQTHLVKLGKAGL-----QF-EGYENLPSDF--NLED 717
DB 595 FFKGHQVASFDQDLTKDKLTVLETLVDLAKKGLETNWAQFPKGAFOYTKDIPAEADK 654
QY 718 LKNIRIKTFLFSQDNF---KLSLDFNNYYDGE- IKAPBFGFLPFLPKELRNRSSNG 772
DB 655 LKFLESK-----KKDPFNQIKIEIHQLSFNILARNDVIKSDGYGVLLLPQSVKTELEGN 709
QY 773 GSQNSNPWEQIISQFKDQNLNSQDLAQASTKIWEKIIIGDENFDO--NRLQYKLLKD 831
DB 710 -----EAQIFBALKKYSLIEN---SAFKTTILDKNLLEGTDPKTFGDFLKAPFLKA 757
QY 832 LQ---ESWINKTRDNLWYTV---LGDKLKVKPKNNLEAKFROISN-LOELLTAFYTS 882
DB 758 AQFNFAFWA-KLDNLQYSPFAIKGETTBEGKREVEVKVKELDNKIKILP--QPPA 814
QY 883 ALSNNWNYQDSGAKSTIIFBEIAR---LDPKVKVKGA----- 918
DB 815 A-----KPEAAKPVAAKPEAAKPVAAKPEAAKPVAAKPEETTKPVATNTNTNGFSLT 866
QY 919 -----DVTQLKPHYAIGFDDNAGKFNQEVIRSSRTIYLTSGSKSLEAD--TIDQLNQAV 972
DB 867 NKPKEDYFPMAFSYKLEYTDE---NKLSTKTPENVFEVLVHQSSEYBEQKIIKELDKTV 922
QY 973 KNAPLGQSFYLDTRERFGVFOKLATS LAVQHKKQEKTLPKLNNDGYTLIHDKLLKVP 1032
DB 923 LNLQYQFQEVKASDQ---YQKLSHPMTTEGSSNGKKAEGAPNQ-----KKAEGAP 972
QY 1033 QISSPEKDMWFEGLKLNQNGSQNVNST-----FGSIESPYFTNFOEDADL 1080
DB 973 SQGKKAEGAPSGK-KAEGTSNQQTTLTNYLPDLGKKIDEIIKQ--GKNWKTEVEL 1029
QY 1081 DDG-----QDSSQGNNSLONQAEAGLLKOKLAILLGNQFYQYQOQNDKEIE 1127
DB 1030 TEDNIAGDAKLYFILRDDSKSG-----DPKKSUKVXITVQKSN-----NQOELK 1075
```

RESULT 14

```
Q8WR55 PRELIMINARY; PRT; 2965 AA.
AC Q8WR55;
ID Q8WR55;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Normocyte-binding protein 1.
GN Name=NBp1;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malayan Camp K-;
```

```
RX MEDLINE=21590493; PubMed=11733572;
RA Rayner J.C., Vargae-Serrato E., Huber C.S., Galinski M.R.,
RA Barnwell J.W.;
RT "A Plasmodium falciparum homologue of Plasmodium vivax reticulocyte
RT binding protein (PvRBP1) defines a trypanin-resistant erythrocyte
RL invasion pathway.";
RL J. Exp. Med. 194:1571-1581 (2001).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=Malayan Camp K-;
RA Rayner J.C., Barnwell J.W., Galinski M.R., Tran T., Huber C.S.,
RA Corcoran V.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF411932; AAL38221.2;
DR ProDom; PD001963; Botulinum; 2.
SQ SEQUENCE 2965 AA; 356914 MW; 98B077462826A8FE CRC64;
```

Query Match 4.1%; Score 395.5; DB 2; Length 2965;
Best Local Similarity 18.7%; Pred. No. 1.7e-07;
Matches 354; Conservative 325; Mismatches 679; Indels 533; Gaps 88;

```
QY 50 IDSVAFKPSIANFT---SDYQSVKKALLNG-----KTFDPKSEFTDFVSKF 93
DB 777 IDDIKIQETLQKQITHIVNNIKTKIKDLKKEFIOHLIKYMNERYQMQGYNLNIYINQY 836
QY 94 DFLTNNGRTVLEIPKQYQVVISFSPEDDKERFLGPHLKEKLEDG---NIAQSAKFI 149
DB 837 EEENNMKQVITTRNIQKIYDNIYAKEKE-IRSGQYKDFITSRKNIYNIENISQNV 895
QY 150 YLLPLDNPKAALGOYSYIVDKNFNNLIHPLSNFSAQSIKPLATRSDDFTAKLNQFNQ 209
DB 896 DMIKNEKK---KIQNCVDK-YSNI-----KOYVMFKNGDTQD-----ENNNNN 937
QY 210 DELWVLEKFDLEALKANRLQATDFSEKGNLVDPFVSFINPQNKQEWASDLNOD-- 268
DB 938 D---IYDKLIVPLDSIQNDIKYNTENF-----ITFNKINTHKK-----NQEM 980
QY 269 -QKTVRLYLRTFSPQAKTILKQYKDETFLLSIDLKASNGTSLFANE--NDLKQDL--- 323
DB 981 MEBFIYVYKR-----LKIFK-----ILNISLKACEKNKSNITNLDKTQELKKI 1024
QY 324 ---DVLDDVSDYFGGQSEIITS-----NSQVKPVSPASERSLKORVKFKDQOKP 370
DB 1025 VTHEIDLL-----QKDLTQISKNKVLNLLNDLKEI---EQYIIDVHKLKKS-- 1071
QY 371 RIEKFSLYEYDALSFYSQ-----LOELVSKPN---SIKDLVNATLARNLRFSLGKYNF 420
DB 1072 --DLFTYEQSKNVFYFKKKNDPDIQKTNKNWEALAKYINEINKYQTLVEKKINV 1129
QY 421 LFDLASHLDYFLVSVKAKTKQSSITKKLFIEPIKLSKSSILGDOEPNITKLFKEVET 480
DB 1130 LLHNSKYVQYFY---DHIINLILQKKNYLENTLTKIQ-----DNEHSYALQOONEEY 1180
QY 481 FKLDNFRD-----VEIEKAFGLLYPGVNELEQA-----RKAQASPEKE 520
DB 1181 QKVNEKDDQNEIKIKOLIEKNKNDILTYENNIEQIEQKIELKTNQAKDDQIVNTLINE 1240
QY 521 KSKKGLKEFSQOKESKAINN--QEG--LEEDDNITERLPENSPYIQOQENAGLGASDPK 578
DB 1241 VKKLIIVYKEVDNQIISNVLKNYEEGKVEYDKNVVQNVNDADDNDIDEINDIDEINDID 1300
QY 579 YM-----IKDVQNRQYLLAKSQIQELIKAKDYTKLAKLSNRHTYNISLR--LKEQLPDV 631
DB 1301 EINDIDEIKOIHDKHFDKDDTKHFDIYHADD-----TRDEYHIALSNYIITELRNI 1351
QY 632 NPR-----IPSSRDIEKAKFVLDDKTEK--NKYW-----QIYSSASP 665
DB 1352 NLQEKNNIKIKFEKESAKHEIKKESQINKETKMDVNVNQLRDIRDQMLDIYKE--- 1408
QY 666 VFQKWSLFGYRYLLGLDPKQTHLVKLGKAGLQFEGYENLPSPFNBLDKNIRIKT 725
DB 1409 -LDEKYSFEN-----KTKIEENINRENINNVIEWYKNIIEYFLAHMNDQDKA 1457
```


QY 561 PIQYQENAGLGASDPKPYMIKDVNQRYYLAKSIOELIKAKDYTKLAKLLSNRHTYNI 620
Db 501 -----NPNYNTQ-----LAKVEEALFKGNKQEIYQALDGNAYAYEF 536
QY 621 SLRLKEQLFDVNPRI--PSSRDIEK-----AKFVLDTKKNKYWQIYSS--ASP 665
Db 537 G-AFKSVLNSWTGKIQHPEKADIQRTHLBOVKGNSVLNQPTTKE-QVSSLSKSN 594
QY 666 VFQKWSLFGYRYLLGLDPKQTHLVLKLGKAGLQEGYENLPSPDNLEDLKN-----721
Db 595 PFKNGHVASFYQDLTKDLTKLVLETLVLAKKGLG-TNWAQPKG-AFYTKDIFABA 652
QY 722 -RIKTPLSQKDNF-----KLSLLDPNNYDGE- IKAPEFGPLFLPKELRRNSNSGGSQ 775
Db 653 DKLFLFLEWKKKDPVNOQINIEHQLSEFNILARNVDIKSDGYGVLLLPQS VKTELEKGN---709
QY 776 NSNSPWEQEIISQFKDQNLNQDLAQSTKIWEKIIGDENEFQ--NNRLQYKLLKDLQ- 833
Db 710 -----EAQIFEALKKYSLIEN---SAFKTTILDKNLEGTDFKTFGDFLKAFFLKAQF 760
QY 834 ---BSWINKTRDNLNWTY-----LGDKLKVKPKNNLEAKFROISN-LOELL-----875
Db 761 NNFAPWA-KLDDNLOYSEAIKGETTKGKREEVDDKVKKELDNKIKGILPQPPAAKPEA 819
QY 876 ---TAFYTSAAALNNWNYQDSGAKSTIIFBEIAE-----LDPKVKKEKVA- 918
Db 820 AKPVAAPVAA-----KPEAAKPVAAKPEAAKPVAAKPEAAKPVAAK 871
QY 919 -----DYYQLKFHYAIGFDDNAGFKNGQEVIRSSRTIYLTSG 956
Db 872 PVATNTNTNTGFSLTWPKEDYFPMAFSYKLEYTDE---NKLSLKTPBINVFLELVH 927
QY 957 KSKLE-ADTIDQLNOAVKNAPLGQSFYLDYTERFGVFOKLATSLAVOHKQEKTLPKKLN 1015
Db 928 QSEVEDQKIIKELDKTVNLNLOYQFQEVKVTSDQ---YQKLSHPMTTEGTQNGKKGSTP 984
QY 1016 NDGYTLIHDKLKVPVPIQISSPEKDWPEGLKQ---NQSQNVNVSTF-----GSI 1064
Db 985 NQG-----KKAEGAPSGKKAEGTNPQGGKAEPTNPQSTSELNLYLPDLGKKIDEI 1037
QY 1065 IESPYFSTNFQEDADLDQD-----QDSSROGNSLNQBEAGLLKQKLAILLG 1112
Db 1038 IKKQ--GKNWKTVEBELIEDNIAGDAKLLYFLIRDSDSKG-----DPKSSLSKVKITVKQS 1090
QY 1113 NQFTQYQQNDKEIE 1127
Db 1091 N-----NNQELK 1097

RESULT 16
O07132
ID O07132 PRELIMINARY; PRT; 1092 AA.
AC O07132;
DT 01-JUL-1997 (TremBLrel. 04, Created)
DT 01-JUL-1997 (TremBLrel. 04, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Adhesin.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J;
RA MEDLINE=98361039; PubMed=9695926;
RA Wilton J.L., Scarman A.L., Walker M.J., Djordjevic S.P.;
RT "Reiterated repeat region variability in the ciliary adhesin gene of
RL Mycoplasma hyopneumoniae."
RL Microbiology 144:1931-1943(1998).
DR EMBL; AF001398; AAC35402.1; -.
DR PIR; T18354; T18354.
DR InterPro; IPR000719; Prot.kinase.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.

SQ SEQUENCE 1092 AA; 123368 MW; 64BFCED1A6657194 CRC64;
Query Match 4.0%; Score 393; DB 2; Length 1092;
Best Local Similarity 20.6%; Pred. No. 5.7e-08;
Matches 261; Conservative 204; Mismatches 464; Indels 340; Gaps 59;
QY 1 MNKKSTLLIATAAIIIGSTVGVGLASKVKYRGVNPQTQGVISQLGLDSDVAFKPSIA 60
Db 1 MSKSKTKFKIGLAGIVGLGVGLTVGLSLAKYRSRPRKIANDFAAKVSTLAFSPYAF 60
QY 61 NFTSDYQSVKALLNGKTFDPKSEFTDFVSKDFPLTNGRTVLEI-----PKYQV 112
Db 61 ETSDYKIVKRWLVDSNN-NIRNKE--KVIDSFSFTKNGDQLEKINFQDPDEYTKAKITF 117
QY 113 VIEFSPEDDKFRPRLGFLHLEKLEDGNIQAQSATKFTYLLPLDMPKAA---LGQSYIVD 169
Db 118 EILEIIPDDVQNQPKVKFQALQKLHNGDIAKSD---IYEQTVAFAKQSNLLVAEFNSLK 174
QY 170 KNFNNL-----IHLPLSNFS---AQSIKPLALTRSSDFIAKLNQFNQDELWYLEKFF 220
Db 175 KITEKLNQOENLSTKITNFADSKTSQKDPSTILRAIDFQYDLNARTNAEDLDIKLANYF 234
QY 221 DLEALKANI-RLQTADFSFEKGNLVDPVYSFIRNPONQKEMASDLNQDQKTVLRYLRT 279
Db 235 PV--LKNLIRNLNAPENKLPNNLGNIFEFSAKD-SSTNQYVSIQNG---IPSLFLKAD 288
QY 280 FSPQAKTILKDYKYKDET--FLSSIDLKASNGTSLFANENDLKDQLDVLDDVSDYFQGQ 337
Db 289 LSQSAREILAS---PDEVQPVINILRMKKDNSSYFLNFDVNNL-----331
QY 338 SETITSNQKVPASERSLKDRVKFKQOQKPRIEKFSLYEY-----DALSFY 386
Db 332 -----TLKN---MQEDLNKAGQNL SAYEFLADIKSGFFPGDKRSSH 370
QY 387 --SQLOELVSKNSIKDLVNATLARNLRFSLGKYNFLPDDL--ASHLDYFLVSKAKIKQ 442
Db 371 TKAEISNLLNKKENIYD-----FGKNGKFNDRNLNSPNLEYSLDAASASLDK 417
QY 443 SSITKFLFELPIKISLSSILGDQ-EPNIK-TLFEKEVTFKLDNFRDVEIEKAFGLLYP 500
Db 418 KD---KSIILIPYRLEIKDKFFADDDLYPDKNLVKEGILKLTGFK----KGPKIDL 469
QY 501 GVNEELEQARKAARASPEKSKGLKBPSCQKEENSKAINNOEGLEEDNITERLSPENS 560
Db 470 NINQQI-----FKTEY---LPFFEKKEEQAKL-----DYGNI- 500
QY 561 PIQYQENAGLGASDPKPYMIKDVNQRYYLAKSIOELIKAKDYTKLAKLLSNRHTYNI 620
Db 501 -----NPNYNTQ-----LAKVEEALFKGNKQEIYQALDGNAYAYEF 536
QY 621 SLRLKEQLFDVNPRI--PSSRDIEK-----AKFVLDTKKNKYWQIYSS--ASP 665
Db 537 G-AFKSVLNSWTGKIQHPEKADIQRTHLBOVKGNSVLNQPTTKE-QVSSLSKSN 594
QY 666 VFQKWSLFGYRYLLGLDPKQTHLVLKLGKAGLQEGYENLPSPDNLEDLKNIKRIKT 725
Db 595 PFKNGHVASFYQDLTKDLTKLVLETLVLAKKGLG-TNWAQPKG-AFYTKDIFABA 652
QY 726 PLSQKDNFK-----LSLLDPNNYDGE- IKAPEFGPLFLPKELRRNS 769
Db 647 DIFAEADKLFLGKKKDPVNOQINIEHQLSEFNILARNVDIKSDGYGVLLLPQS VKTELE 706
QY 770 NSGGSQNSPWEQEIISQFKDQNLNQDLAQSTKIWEKIIGDENEFQ--NNRLQYKLL 828
Db 707 GKN-----EAQIFEALKKYSLIEN---SAFKTTILDKNLEGTDFKTFGDFLKAFF 754
QY 829 LKDLQ-----ESWINKTRDNLNWTY-----LGDKLKVKPKNNLEAKFROISN-LOELLTAFY 879
Db 755 LKAAQFNFPWA-KLDDNLOYSEAIKGETTKGKREEVDDKVKKELDNKIKGILP--Q 811
QY 880 TSAALSNNWNYQDSGAKSTIIFBEIAELD-----PKVKEKVA- 918
Db 812 PPAA-----KPEAAKPVAAKPEAKPETTKPVAAKPVAAKPVAAKPVATNTN 863

QY 919 -----DVYQLKFHYAIGFDDNAGKFNQEVIRSSRTIYLTSGKSKLEAD-T 964
Db 864 TWTGSLTNKPKEDYFPMAFSYKLEYTDS-----NKLSTKPTINVFLELVHQSEYEEQKI 919
QY 965 IDQLNOAVKNAPLGLQSFYLDTERFVGFOKLATSLAVQHKQEKTLPKKLNNDGYTLIHD 1024
Db 920 IKELDKTVLNLQYQFQVKVTSEQ---YQKLSHPMTTEGSPNQGKKAEGAPNQG----- 970
QY 1025 KKKPV-----IPQISSSPKDMFEGKLNQ-----NGSQNNVNSTF-----G 1062
Db 971 --KKAEGAPSGQKKAEGAPNQGKKAEGPSQKKAEGASNOQSTTTLTNLYPELGKKID 1028
QY 1063 SIIESPYFTNFOEDADLDQD-----QDSSQGN-----NSLDN 1097
Db 1029 BIIKKQ--GKNWTEVELIEDNIADAKLLYFVLRRDSSKSGDPKSSVKVKITVQSNNN 1086
QY 1098 QEAGLLKQK 1106
Db 1087 QE---LKSK 1092

RESULT 17
Q6R5H7
ID Q6R5H7 PRELIMINARY; PRT; 1092 AA.
AC Q6R5H7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Adhesin.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID:2099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25934;
RA Boisvert A., Cheikh Saad Bouh K., Khevar A., Shareck F., Dea S.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY512905; AAS00640.1; -
DR InterPro; IPR000719; Prot kinase.
DR PROSITE; PS00107; PROTEIN KINASE ATP; UNKNOWN 1.
SQ SEQUENCE 1092 AA; 123324 MW; 624518A7ACA3DB54 CRC64;

Query Match 4.0%; Score 391.5; DB 2; Length 1092;
Best Local Similarity 20.3%; Pred. No. 6.5e-08;
Matches 259; Conservative 209; Mismatches 473; Indels 335; Gaps 58;

QY 1 MKNKKSTLLATAAIIIGSTVGTGVLGASKVKYRGVNPQTQVISQLGLIDSVAPKPSIA 60
Db 1 MSKKSTFKIGLTAGIVGLGVGLTVGLSSLAKYRSESPRKANDPAAKVSTLAFSPYAP 60
QY 61 NFTSDYQSVKALLNGKTFDPKSEFTDFVSKPFDLTNNRTVLEI-----PKYQV 112
Db 61 ETDSYKIVKRWLVDSNN-NIRKE--KVIDSFPFTKNGDQLEKINFQDPYTKAKITF 117
QY 113 VISEFSPDDKERFRLGHLEKLEDCNIAQSTKFIYLLPLDMPKAA---LQOYIYVD 169
Db 118 EILEIIPDDVNQNFVKFQALQKHLNGDIKASD---IYEQTVAFAQSNLLVAEFNFSLK 174
QY 170 KNFNLI-----IIHPLSNFS---AQSIKPLATRSDDFIKLNQFNNOELWVLEKFF 220
Db 175 KITEKLNQJENIESTKJINFADEKTSQKDPSTLRAIDFOYDLNTARNADLDIKLANYF 234
QY 221 DLBAKANI--RLQTAQDFEFKGNLVPFVYFIRNPNQKEMASDLNQDKTVRLYRTE 279
Db 235 PV--LKNLINLNANPENLNNLGNIFEFSFAKD--SSTNQYVSIQNG---IPSLFLKAD 288
QY 280 FSPQAKTILKDYKKBOT--FLSIDLKASNGTSLFANENDLKQDLVDLLDVSDFGGQ 337
Db 289 LSQSAREILAS---PDSEVQVINILRLMKKNDSSYFLNFEDFYNNL----- 331
QY 338 SEITITSNQVKVPASERSLKDRVKFKQKQKPRIEKFSLEY-----DALSFY 386

Db 332 -----TLKN---MQKEDLNAGQNLSAVEFLADIKSGFPDGPGRSSH 370
QY 387 --SOLQELVSPNSIKDLVNTATLARNLRFSLGKTNFLFDDL--ASHLDYYFLVSKAKIKQ 442
Db 371 TKABISNLLNKENIYD-----FGKYNKPNDRNLNFPNLEYSYDAASASLDK 417
QY 443 SSITKCLPIELPIKISLSSILGDQ-BPNIK-TLPEKEVTFKLNFRDVEIEKAFGLLYP 500
Db 418 KD---KSTILPYRLEIKDKFFADLDYPTDNLIVKESILKLTGFK-----KGPKIDLP 469
QY 501 GVNBELEOARQAASPEKESKGLKEFSQKENSKAINNQGLEBDDHITERLPENS 560
Db 470 NINQOI-----FKTEY---LPFFEKGEEQAL-----DYGNI----- 500
QY 561 PIQOENAGLIGASPDKPYMIKDVQNRQYIYLAQSOIELIKAKDYTKLAKLLSNRHTYNI 620
Db 501 -----NPNTQ-----LAKVEEALFKGNKQEIYQALDGNAYEF 536
QY 621 SLRLKEQLFDVNPRI--PSSRDIEK-----AKFVLDKTEKNKYWQIYSS--ASP 665
Db 537 G-AFKSVLNSWTGKIQHPEKADIQRFTHLBOVLKGSNSVLNQPTTKE-QVSSLSKSN 594
QY 666 VFQNKWSLFGYRYLLGLDPKQTIHELKVLGOKAGLQEGYENLPSDNLELDKNIRKT 725
Db 595 PFKNGHVASVFQDLLTKDKLTLETLVDLAKKWL-----ETNRAQFPKEVQYTK--- 646
QY 726 PLFSQKDNFK-----LSLLDFNNYDGE--IKAPEGLPLFLPKELRRNSS 769
Db 647 DIFABADKLFLEGKKDPYNOIKHQLSPNILARNDVIKSDGYGVLLLPQSVKTELE 706
QY 770 NSGSGSONSPWEQIIISQFKDQNLSDQLAQFSTKIWEKIIGDENEFDQ--NNRLQYL 828
Db 707 GKN-----EAQIFALKKYSIJEN--SAFKTTILDKNLLSGTDFKTFGDFLKAFP 754
QY 829 LKDLQ-----ESWINKTRNLNWTY-----LGDKLKVKPKNNLEAKPRQISN-LQELLTAFY 879
Db 755 LKAAQFNNFAPWA-KLDDNLQYSFEAIKKGHTTKGKEEVDKKVKELDNKIKGILP--Q 811
QY 880 TSAALSNNWNYQDSGAKSTIIFERIAELD-----PKVKEKYGA----- 918
Db 812 PPAA-----KPEAAKPVAAKPEAAKPEETKPVAAKPEAAKPVAAKPVATNTN 863
QY 919 -----DVYQLKFHYAIGFDDNAGKFNQEVIRSSRTIYLTSGKSKLEAD-T 964
Db 864 TWTGSLTNKPKEDYFPMAFSYKLEYTDS-----NKLSTKPTINVFLELVHQSEYEEQKI 919
QY 965 IDQLNOAVKNAPLGLQSFYLDTERFVGFOKLATSLAVQHKQEKTLPKKLNNDGYTLIHD 1024
Db 920 IKELDKTVLNLQYQFQVKVTSEQ---YQKLSHPMTTEGSPNQGKKAEGAPNQG----- 970
QY 1025 KKKPVIPQISSSPKDMFEGKLNQ-----NGSQNNVNSTF-----GS 1063
Db 971 --KKAEGAPSGQKKAEGAPNQGKKAEGPSQKKAEGASNOQSTTTLTNLYPELGKKIDE 1029
QY 1064 IYESPVFTNFOEDADLDQD-----QDSSQGNNSLDNQEAGLLKQKLAIDL 1111
Db 1030 IIKKO--GKNWTEVELIEDNIADAKLLYFVLRRDSSKSG-----DPKSSLSKVKITVQ 1082
QY 1112 GNQFIQYQOONDKEIB 1127
Db 1083 SN-----NNQELK 1090

RESULT 18
Q49542
ID Q49542 PRELIMINARY; PRT; 1108 AA.
AC Q49542;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE P97.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

```
OX NCBI_TaxID=2099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=232;
RX MEDLINE=97175562; PubMed=9023217;
RA Hsu T., Artushin S., Minton F.C.;
RT "Cloning and functional analysis of the p97 swine ciliium adhesin gene
RL of Mycoplasma hyopneumoniae.";
RL J. Bacteriol. 179:1317-1323(1997).
DR EMBL; U0901; AAB47806.1; -
DR PIR; T18353; T18353.
DR InterPro; IPR000719; Prot_kinase.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN 1.
SQ SEQUENCE 1108 AA; 124903 MW; 239CF62D61E1FE7 CRC64;

Query Match 4.0%; Score 391.5; DB 2; Length 1108;
Best Local Similarity 20.6%; Pred. No. 6.6e-08;
Matches 258; Conservative 210; Mismatches 486; Indels 325; Gaps 57;

QY 1 MNKKSTLLATAAAIGSTVFTVGLASVKYRGVNPQTQGVISQGLIDSVAFKPSIA 60
DB 1 MSKSKTFKIGLTAGIVGLGVGLTVGLSLAKYRSSEPRKIANDFAAKVSTLAFSPYAF 60
QY 61 NFTSDYQSVKALLNGKTFDPKSEFTDFVSKFDFLTNGRTVLEI-----PKYQV 112
DB 61 ETDSDYKIVKRWLVDSNN--NIRNKE--KVIDSFSFFTKNGDQLEKINFQDPYTKAKITF 117
QY 113 VISFSPEDDKERPRGLPHLKEKLEDGNIAGSATAKFIYLLPLDMPKAA---LGQYSYVD 169
DB 118 EILEIIPDVNQNFVKVQALQKUNGDIASD---IYEQTVAFAKQSNLLVAEFNFSLK 174
QY 170 KNFNNL-----IIHPLSNFS---AQSIKPIALTRSSDFIAKLNQFNQDELWYLEKFF 220
DB 175 KITEKLNQNIENLSTKINTFADEKTSQKDPSTLRAIDFQYDLNATARNPDLDIKLANYP 234
QY 221 DLEALKANI--RLQTADFSEKGNLVDPPVYFIRNPQNKQEWASDLNQDQKTVLYLRT 279
DB 235 PV--LKNLINLNAPENKLPNNLGNIFEFSKAD--SSTNQYVSQIQO---IPSLFLKAD 288
QY 280 FSPQAKTILKDYKYKDET--FLSSIDLKASNGTSLFANENDLKQOLDVLDLVDSDYFGG 337
DB 289 LSQSAREILAS---PDEVQPVINILRLMKDQNSSYFLNFDFVNNL-----331
QY 338 SETITSNQVVPASERSLKDVRVKFKDQKQKRIEKFSLVEY-----DALSFY 386
DB 332 -----TLKN--MQKEDLNAGQNLSAVEFLADIKSGFPFGDKRSSH 370
QY 387 --SQQLVSKPNSIKDLVNATLARNLRFSLGKYNFLPDDL--ASHLDYFLVSKAKIKQ 442
DB 371 TKAEISNLLNKKENIYD-----FGKYNKGFNDRLNSPNLEYSLDAASALDK 417
QY 443 SSITKFLFELPIKISLSSILGQ--EPNIK--TLFEKEVTFKLNFRDVEIEKAFGLLYP 500
DB 418 KD---KSIVLIPYRLIEIKDFADDDLYDPTKONILVKEGILKLTGFK-----KGSKIDLP 469
QY 501 GVNELEQAKQAQASFEKSKKGLKEFQSKQKENSKAINNQGLBEDDNIITRLPENS 560
DB 470 NINQOI-----FKTEY---LPPFEKGEQAKL-----DYGNIL-----500
QY 561 PIQVQENAGLGASPKDKYMLKQVQNRQYLLAKSQIQELIKAKDYTKLAKLLSNRHTYNI 620
DB 501 -----NPNYNTQ-----LAKVEEALFQGNKQNKQYIQLADGNAYAYEF 536
QY 621 SLRLKEQLFDVNPRI--PSSRDIEK-----AKFVLDTKTEKNKYQIYSS--ASP 665
DB 537 G-AFKSVLNSWTGKIQHEKADIQFTRHLQVKGISNSVLNQPTTKE-QVISELKSNN 594
QY 666 VFQNKWSLFGYYRYLLGLDPKQTHIELVKLQKAGL-----QF-EGYENLSPDF--NLED 717
DB 595 FFKNGHQVASYFDQLLTTKDKLTILETLVDLAKKWLGTETNRAQFPKGVQYTKDFAEADK 654
QY 718 LKNIRIKTFLPSQKNF---KLSLLDFNNYVDGE--IKAPEGLPLFLPKELRRNSSNG 772
```

```
DB 655 LKFELEK-----KKDPYNOIKETIHQLSFNILARNVDVIKSDGFYGVLLLPQSIVTELEGKN 709
QY 773 GSONSNPWEQEIIISOQKDNLSNQDLAQFSFKIWEKIIGDENEFDQ--NNRLQYKLLKD 831
DB 710 -----EAQIFEAALKVYSLEIEN---SAFKTTILDKNLLGCTDFKTFGDFLKAPFLKA 757
QY 832 LQ-----ESWINKTRDNLWYTY-----LGDKLKVKPKNNLEAKFRQISN-----870
DB 758 AQFNNEAPWA-KLDDNLOYSFEAIKGETTKEGKREBEVDKKVKELDNKIKGILPQPPAAK 816
QY 871 ---LQELLTAFTYTSAAALSNMNNVYQDSGAKSTIIFBEIAB-----LDPKVKE 914
DB 817 PEAAKPVAAKPEITKPVAAKPEAAKPEAAKPVAAKPEAAKPVAAKPEAAKPEAAK 876
QY 915 KVGA-----DVYQLKPHYAIGFDDNAGKFNQEVIRSSRTIYL 952
DB 877 PVAAPKPEAAKPVATNTGFSLTNPKPKEDYFPMAFSYKLEYTDE---NKLSLKTPINVFL 932
QY 953 KTSQSKGL-EADTIDQLNQAVKNAPIGLQSFYLDTERFGVFOKLATSLAVQHKQKEKTL 1011
DB 933 ELVHQSEYEQEIIKELDKTVLNLOQFOEVKVTSDQ---YQLSHPMMTTEGSSNQGKKS 989
QY 1012 KLLNNDGYTLIHDKLKKPVIPQISSSPKDWFEGLNQNGSQSNVNV-----T 1060
DB 990 EGTPNQ-----KKAEGAPNQGKAEGTPNQGKKAEGAPSQOSPTELNTYLPDLGKK 1042
QY 1061 FGSIIESPYSFTNFQBDADLDQDG-----QDSRQGNNSLDNOEAGLLKOKLA 1108
DB 1043 IDEIHKQ--GKNWKTVEVELIEDNIAGDAKLFLYLRDSDSKG-----DPKSSLSVKKIT 1095
QY 1109 ILLGNQFIQYQONDKEIE 1127
DB 1096 VKQSN-----NNOEPE 1106

RESULT 19
Q6RS16 PRELIMINARY; PRT; 1098 AA.
AC Q6RS16;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Adhesin.
OC Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2099;
RN [1]
RP SEQUENCE FROM N.A.
RA Boisvert A., Cheikh Saad Bouh K., Kheyar A., Shareck F., Dea S.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY512896; AAS00631.1; -
DR EMBL; AY512897; AAS00632.1; -
DR EMBL; AY512899; AAS00634.1; -
DR EMBL; AY512901; AAS00636.1; -
DR EMBL; AY512895; AAS00630.1; -
DR InterPro; IPR000719; Prot_kinase.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN 1.
SQ SEQUENCE 1098 AA; 123873 MW; DC259FC350FBF14A CRC64;

Query Match 4.0%; Score 390.5; DB 2; Length 1098;
Best Local Similarity 20.6%; Pred. No. 7.2e-08;
Matches 265; Conservative 202; Mismatches 449; Indels 371; Gaps 60;

QY 1 MNKKSTLLATAAAIGSTVFTVGLASVKYRGVNPQTQGVISQGLIDSVAFKPSIA 60
DB 1 MSKSKTFKIGLTAGIVGLGVGLTVGLSLAKYRSSEPRKIANDFAAKVSTLAFSPYAF 60
QY 61 NFTSDYQSVKALLNGKTFDPKSEFTDFVSKFDFLTNGRTVLEI-----PKYQV 112
DB 61 ETDSDYKIVKRWLVDSNN--NIRNKE--KVIDSFSFFTKNGDQLEKINFQDPYTKAKITF 117
QY 113 VISFSPEDDKERPRGLPHLKEKLEDGNIAGSATAKFIYLLPLDMPKAA---LGQYSYVD 169
```

Db 118 EILEIIPDDVNQNFVKVFOALQKLNHDIKSD-----IYEQTVAFAKQSNLLVAEFNSLK 174
Qy 170 KNFNML-----IIHPLSNFS---AQSIKPLALTRSSDFIAKLQFNQDELWVYLEKFF 220
Db 175 KITEKLNQIENLSKTIITFADEKTSQKDPSTLRAIDFOYDLNTARNAEDLDKLANVF 234
Qy 221 DJEALKANI-RLQADFSPEKGNLVDPPVYFIRPNQKQEWASDLNQDQKTVRLYLRT 279
Db 235 PV--LKNLNLNAPENKLPNINIFEFSPAKD--SSTNQVSVTQNG---IPSLFLKAD 288
Qy 280 FSPQAKTILKDYKDET--FLSSIDLKASNGTSLFANENDLKQDLDDVLLDVSDYFGQ 337
Db 289 LQSAKEILAS---PDEVOPVINILRMKONSSVFLNPFEDFVNL-----331
Qy 338 SETITNSQVPPASERSLDRVRPKDQKPRIEKFSLEYE-----DALSPY 386
Db 332 -----TLKN--MQKEDLNAKGQNLISAYEFLLADIKSGFFPGDKRSSH 370
Qy 387 --SQQLVSKNSIKOLVNATLARNLRFSLGKYNFLFDDL--ASHLDYFVLVSKAKIKQ 442
Db 371 TKSEISNLNKKENIYD-----FGKYGKFNDRLNSPNLEYSLSLDAASGLDK 417
Qy 443 SSITTKLFIELPKISLKSILGQD-EPNIK-TLPEKEVTFKLDNFRDVEIEKAFGLXP 500
Db 418 KD---KSIILIPVLEIKDKFPADLDYPTKNILVKEGILKLTGPK-----KGPKIDL 469
Qy 501 GYNEELEQAKQARASFEKSKGKGLFEFSQOKEENSKAINNQGLEEDDNTERTLPENS 560
Db 470 NINQI-----FKTEY---LPFEKGEQAKL-----DYGNIL-----500
Qy 561 PIQYOENAGLGSPPKPMKDVQNRYYLAKSOIQELIKADYTKLAKLNSNRHTYNI 620
Db 501 -----NPYNTQ-----LAKVEEALFKGNKQEIYQALDGNVAYEF 536
Qy 621 SLRLKQLFDVNPRI--PSSRDIEK-----AKFVLDKTEKNYQWYSS--ASP 665
Db 537 G-AFKSVLNSWTGKIQHPKADIQTRHLLEQVKGNSVNLQNPQTKB-QVSSLSKSN 594
Qy 666 VFQNKWSLFGYRYRLGLDLPKQTIHBLVKGQKAGL-----QF--EGYENLSPDFNLEDL 718
Db 595 PFKNGHQVASYFDLLTKDKLTVLELYDLAKKWLGTNAQAPQFEAFQVTKDIPAEAD- 653
Qy 719 KNTRIKTPLFSQKNF-----KLSLDFNNYDGE-IAKPEGLPLPLPKELRNSNSGG 773
Db 654 ---KLKFLWKKKDPYNQIKHQLSFNILARNDVIKSDGFYGVLLLPQSVTELEGKN- 709
Qy 774 SQNSNPWEQEIISQFKDQNLNQDLAQFSTKIWEKIIGDENEFDQ-NNRLOQYKLLKDL 832
Db 710 -----EAQIFEAALKYSLIEN---SAPKTTILDKNLLEGTDFKTGDFLKAPFLKAA 758
Qy 833 Q-----BSWINKTRDNLTYTY---LGDKLKVKPKNNLEAKFROISN-LQELL-----875
Db 759 QFNFPAPWA-KLDNLQVSEAIKKGTTKEGKREEVKVKKELDNKIKGILPOPPAAKP 817
Qy 876 -----TAFYTSAAALSNNWNYQDSCAKSTIIFEEI 905
Db 818 EAAKPVAAKPEAAKPVAAKPEAAKPVAA-----KPEAAKPVAAKPEA 869
Qy 906 AE---LDPKVKVGA-----DVYOLKPHYAIGFDDNA 935
Db 870 AKPVAAKPEAAKPVAAKPVAAKPVAAKPVAAKPVAAKPVAAKPVAAKPVAAKPVAA 928
Qy 936 GKFNQEVIRSSRTIYLTSGSKLEAD-TIDQLNAVKNAPLQSGFYLDTFRFGVQK 994
Db 929 ---NKLSLKTPENVFLVHQSEYEQKIKELDKTVLN-----965
Qy 995 LATSLAVQHQKQKTLPKKLNNGYTLIHDKLKPVIPOISSPEKDWFGKLNQ-----N 1050
Db 966 -----LQYQFQE-----VKVASDQY-----QKLSHPMMTEGSSNQKK-AEGAPNQKKAE 1010
Qy 1051 GQSONVNVST-----FGSIIETSPYSTNFQEDADLDQD-----QD 1086
Db 1011 GTSNQOQNTTTELNYLPKELGKKIDELIKKO--GKNWTEVELIEDNAGDAKLLYFVLRD 1068

Qy 1087 DSRQGNNSLDNQEAGLLKQKLAILLGN 1113
Db 1069 DSKSG-----DPKSSSLKVKITVQSN 1090
RESULT 20
QTRFS2
ID QTRFS2 PRELIMINARY; PRT: 2740 AA.
AC QTRFS2;
DT 01-MAR-2004 (TREMELrel. 26, Created)
DT 01-MAR-2004 (TREMELrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE 235 kDa thoptry protein.
GN Name=PY04630;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perte M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519 (2002).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01001422; EAA16521.1; -.
DR HSSP; P03069; 1GCM.
DR InterPro; IPR006499; ReticulocyteBP.
DR TIGRFAMs; TIGR01612; 235kDa-fam; 1.
SQ SEQUENCE 2740 AA; 321463 MW; 61A4058FF525E599 CRC64;
Query Match 4.0%; Score 390; DB 2; Length 2740;
Best Local Similarity 18.8%; Pred. No. 1.5e-07;
Matches 428; Conservative 324; Mismatches 775; Indels 748; Gaps 102;
Qy 39 PTOGVISQLGL-----IDSVAFKPSIANPTDQYQSVKALLNGKTPDPKSSBFTDFVSKF 93
Db 520 PFRSII---GFVEWRINQTLTEVEAGVKASYESVKWRKSMVEINKLEENKVIKL 576
Qy 94 DFLTN-----GRVLEIPKAYQVVISPEFSDDKERFLRGLPHLKEKLED 138
Db 577 ETQINDLFNQYLKINDENIYLNKLKLEKELKNI-----SDKNEYVVKAVDLKKTEN 630
Qy 139 GMI-AQSATKFIYLLPLDP-----KALAQ-YSVIVDKNFNNLIHPLSNF 183
Db 631 NNIIYDELTK---TSPYQVPEHLKNTDTIYNTIKLELSQIYEDDIDKYNEL-----679
Qy 184 SAQSIKPLALTRSSDFIAKLQFNQDELWVYLEKFPDLEALKANI--RLQADFSFEKG 241
Db 680 -----SSIVOENDIDNVED---KTKLDDLOSKIDNVYSKIQ-----712
Qy 242 NLVDPFVYFIRPNQKQEWASDLNQDQKTVRLYLRTFESPOAKTILKDYKVKDETFLSS 301
Db 713 NMENETVESHLTNIETNKNLSD---TILAIKKYIGEISKDLNKTLEDFOKKEK-----764
Qy 302 IDLKASGTSFLANENDLKQDLDDVLLDVSDYFGQ-----SETITNSQVXP 349
Db 765 ---ELSNKINDYAKENDQLNVYKSKISEIRNHYSQINIDNTKEGEAKQNYDKSNEHMTK 821
Qy 350 VPASERSLK--DRVFKKDKQKPRIEKF-----SLYEYDALSFYSQQLVSKP-----396

"A Plasmodium falciparum homologue of Plasmodium vivax reticulocyte binding protein (PvRBP1) defines a trypsin-resistant erythrocyte invasion pathway." J. Exp. Med. 194:1571-1581 (2001).

[2]

SEQUENCE FROM N.A.

RC STRAIN=768;

RA Rayner J.C., Barnwell J.W., Galinski M.R., Tran T., Huber C.S.,

RA Corridor V.;

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF411933; AAL38222.2; -

DR ProDom: PD001963; Bofulinum; 2;

SQ SEQUENCE 2976 AA; 358102 MW; C00P4D4A1556867C CRC64;

Query Match

Best Local Similarity 18.6%; Pred. No. 2.8e-07;

Matches 352; Conservative 325; Mismatches 684; Indels 532; Gaps 87;

50 IDSVAFKPSIAFT---SDYQVKKALLG-----KTFDPKSSSEFTDFVSKF 93

779 IDDIKIQTQLQITHVNNIKTKDKLKEFTQHLIKYNNERYQNNQGYNNLTNYINQY 838

94 DFLTNGRTVLEIPKQYQVVFSEFSDDEKFRFLGFLHKEKLEDG---NIAQSAKTFPI 149

839 EENNMMKQYITIRNIQIYDNIYAKKE-IRSGQYKDFITSRKNIYINREINISKV 897

150 YLLPLDMPKAALQGVSYIVDKFNPNLIIHPLSNFSAQSIKPLALTRSSDFIAKLNQFNQ 209

898 DMKNEEK---KIQNCVDK-YNSI-----KQCVKMLKNGDQDQNNNNNNNN 943

210 DELWVLEKFDLEALKANIRLOTADFPEKGNLVDVPFYVIRPQONKEWASDLNQDQ 269

944 ND--IYDKLIVPLDSIKQIDKYNTEHNF-----ITFNKINTHNKK-----NQB- 986

270 KTVRLVLTFRFSQAQITLKQYKDETFLS--SIDLKASNGTSIPANE-NDLKQDL--- 323

987 -----MMEEFI-----YAYKELKILKILNISLKACENKNSINTLNDKTQELKKI 1031

324 ---DVDLLDVSDFGQSETITS-----NSQKVPVPASERSLKRVKPKDQKQKP 370

1032 VTHEIDLL-----QKDLTSQISQNNVLLNDLLKEI---EQVIIDVHLKXKSN-- 1078

371 RIEKFSLEYDALSFSQ-----LOELVSKN---SIKDLVNATLARNLFSGLKYNF 420

1079 --DLFTYEQSKNYFYFKNKONCFDQTKTNKNWEALIKYINELKNKYQTYLKKXINV 1136

421 LPDDLASHLDYVFLVSKAKIKQSSITKGLFIELPIKISLSSILGQEPNITLPEKVT 480

1137 LHNSKSYVOYFY---DHIINILQKNVLENTLTKTIQ-----DNEHSLVALQNEEY 1187

481 FKLDNFRD-----VBEIKAFGLLYPGVNEELEQA-----RKAQASFEKE 520

1188 QKVNEKDQNEIKKIKQIEKNKDILTYENNIEQIEQKNIELKTAQNKDDQIVNTLNE 1247

521 KSKKGLKFSQOEKNSKAINN-QEG-LEEDDNIITELPENSPIQYQOENAGLSPDKP 578

1248 VKKKIITYTEKVDNQISNLKNEYEGKVEYDKVQNVQVNDADPTNDIDEINDIDEIND 1307

579 YM-----IKDVQORYLAKSQIEILKAKDYTKLAKLLSNRHYNISLR--LKEQLFDV 631

1308 EINDIDEIKDHIKHFDQTHFDYHADD-----TRDEYHALSNYIKTELNI 1358

632 NPRISSRDIEKARFVLDKTERKNYQWYSSASPVFNQKWSLFGYRYLLGLDPKQTIHE 691

1359 N-----LOEIKNNIIKIFKEFKSAHKEIKKESQINKKEFTKMDVWINQLRDID-RQMLDL 1412

692 LVKLGQAGLQFEGYENLPDFN---LEDLKNIR---IKTPLSQDNKPKSLLDNNYY 745

1413 YKELDEKY-----SEFNKTKIEEINNRENNINNVIEIWKYKNIIEFYLRHNDQK 1461

746 DGEIKAFEGFLPLF-----LPKELR-----RNSNSN-----G 772

1462 DKAAKYME-NIDTYKNNIEIISKQINPENTVETLNKSNMYSYVEKANDLFYKQINNIIN 1520

773 GSQNSNRPW---EQEIIISQFKDQNLNSQDOLAQFSTKIWEKIIGDENEFDQ---NNRL- 824

1521 SNQLKNEAFTIDELQNTQKRNLLTKQOIQTNEI-----ENFIEIKININILV 1573

825 --QYK-LLKOLQESWINKTRDNLWYTVLGDRLKVKPKNNLEAKFRQISNLQELTAFYTS 881

1574 LTNYSILQDISQ--NINHVSITYEQL-----HNLVIKLEE--EKEQMTLYHKS 1619

882 AALNNNNYODSAGAKSTII--FEEIAELDPKVEKVGADYVQLKPHYAIGFDDNAGKNQ 940

1620 NVLHNQINFEDAFINLLINIEKINDITHKEK--TNIYIMDVNKS---KNNALQYFPH 1674

941 EVIRSSRTIYLTSGSKLEADTTIDOLNOAVKNAPIGLQSPFYLDTERFGVFOKLATSLA 1000

1675 NTLRGNEKIEYLLKLNKSTNQOITLOELKQVQENVE-KVQDIYNOCTIKY----- 1722

1001 VQHKQKEXTLPKGLN--NDGYTLIHDKLKPKVPIQISSSPKDMFEGKLNQNGQSON--- 1055

1723 -----BEKIKKNYHIITDYNKINDILHNSFIQINME-----SSNNKKQTKQI 1766

1056 ---VNVSTFGSIIESPSTNPOEDAD-----LDQGDQ---DSROGNSLNDQEA 1100

1767 IDIINDKTFFEHKTSKTKINMLKEQSQMKIDKTLTNEQALKLFVDINSTNNLNDMLS 1826

1101 GLLKQKLAILLGNQFIQYVQONDKEIB--PEINVEKVEL----- 1139

1827 EINS-----TONNIHTYIOBANKSFDPKIICDQNVNDLLNKLGLDNLNHNHLKNLON 1880

1140 ---SPRVEFKL--AKTLEDNGKTIIRVLSDETSLIVNTTIEKTPMESAPEVEDTKWV 1192

1881 EIRNNMLEKNPMLDKSKKIDSEKKLDILKYNISN--INNSLDLKKY--VEEALFQKVK 1936

1193 EYDPRTPPLAATKFKVLK-----FKQIIPVDGS-----GNISD 1225

1937 EKAENQENIEKIOEINTLSDVPKPPFFQTLNTDSSQHEKDINNVTYKNNIDEIYN 1996

1226 KWLASIPILVHQMLRLSPVW---KTRHGLKTEQOQQOQQOQQOQQOPOK----- 1273

1997 VFIQYNNLIQYSSIFASTLNYIOTKEIKESIK--EQNLQNEKEASVLLKNIKINET 2055

1274 ---KAVRKEELETYNPKDEFNINPLTPKAHRLTSLN-----VNNDPNYKIEDLKVI 1323

2056 IKLFOIKNERQNDVHNKEDVNLQVLYNVMKNEBQKKYKNDVHMDKNY-----V 2108

1324 KNEAGDQ-----LAFSLRANNIKRLMNTPIITADYNPPFYNNEDWRSIDKLYNNKGN 1376

2109 ENNNGEKEKLLKETISSYVDKINNINNKL-----YIYKKE-----DTYFNMMIK 2153

1377 VS-----SHQQAAGNQGSGLIIRLNKNIKPTPTPALIALKDRNNTLSNY 1424

2154 VSEIINIIKKQOQNEQRIVINAEVSSLINK--DSEIKKE--INNQIIEI--KHNENISNI 2210

1425 SKIIMIKPKYLVERSIGVPMWSTGLDYGTSQTKDGTSSSSQKQGFQDQFQALGLKNT 1484

2211 FKDIQNIKKQ-----SQDIITNM-----N 2229

1485 EYHGKLGSLIRFPDQNE-----LAKIDASNKKGE--EKLLK----- 1520

2230 DMHKSTILLVDIIOKKEEALNKQKILRNIDNLNKKENIIDKVIKNCDDDYKDILIQNE 2289

1521 -SYDLPKVNLNVEKSKPKIAKGTWNIHPDQKEYPNPKLPENY-----LNLVLNQPW 1573

2290 TEYQKLQNLNHTYEKKKSI-----DILKIKNIKQNTQYKKNKLEQMTYINQSI 2340

1574 KYTLVNSSDFITNLFVEPEGSDRGSTKLQVQI-----KQVNNYADWGSAY--ITFWY 1626

2341 EQHVFINADILQN-----BKILEEIIKNLDILDEQIMTYHNSIDELYKLGIOQ 2389

1627 DKNITNQNVITANIADVFIKDVKELENDTKL 1659

2390 DNHLIITTSVAVVNNKNTKIMIHKKQKEDIQKI 2422

[1]
SEQUENCE FROM N.A.
RA Boisvert A., Cheikh Saad Bouh K., Khevar A., Shareck F., Dea S.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
RM EMBL; AY512904; AAS00639.1; -
DR InterPro; IPR000719; Prot kinase.
DR PROSITE; PS00107; PROTEIN_KINASE ATP; UNKNOWN 1.
SQ SEQUENCE 1093 AA; 123337 MW; 3F883FB0B2E16AFO CRC64;

Query Match 4.0%; Score 389; DB 2; Length 1093;
Best Local Similarity 20.6%; Pred. No. 8.3e-08;
Matches 262; Conservative 210; Mismatches 476; Indels 322; Gaps 59;

Qy 1 MKNKKSTLLLATAAAIIGSTVFTGVYGLASKYKVRGVNPTQGVISQLGLDSDVAFKPSIA 60
Db 1 MSKSKSTFKTGLTAGIVGLGVFTGLVGLSLAKYRSRSPRKIANDFAAKVSTLAFSPYAF 60
Qy 61 NFTSDYQSVKALLNGKTFDPKSEFTDFVYSKDFLTNNGRTVLLEI-----PKYQV 112
Db 61 ETUSDYKIVRWLVDSNN-NIRNKE--KVIDSFSFFTKNGDQLEKINFDPQPYTKAITF 117
Qy 113 VISEFSPEDDKERFLRGFLHKLKLEKLDGNIAQASATKFIYLLPLDMPKAA----LGQYSYIVD 169
Db 118 EILIEIIPDDVNVQFKVKFQALQKLHNGDIAKSD---IYEQTVAFAKQSNLLVAEENFSLK 174
Qy 170 KNFNNL-----IHLPSNFS---AQSIKPLALTRSSDFIAKLQNFQNDQELWVYLEKFF 220
Db 175 KITEKLNQOQIENLSTKITNFADKETSQKDPSTLRAIDFQYDLNTARNASDLDLQKANYF 234
Qy 221 DLEALKANI-RLQTADFSFEKGNLVDPFVYSFTIRNPONQEKWASDLNQDKTQVRLYLIRTE 279
Db 235 PV--LKNLIRNLNAPENKLPNNLGNIFESPAKD--SSTNQYVSIGNQ---IPLFLKAD 288
Qy 280 FSPQAKTILKDYKYKDT--FLSISDLKASNGTSLFANENDLKQDLVDLLDVSVDYEGGQ 337
Db 289 LSQAREILAS--PDSEVQPVINILRLMKDNGSYFLNFDVFNVL----- 331
Qy 338 SETITSNQVQVPASERSLKDRVKPKDQKPKRIEFSLEY-----DALSFY 386
Db 332 -----TLKN---MQKEDLNAGQNLSAYEFLLADIKSGFFPGDGKSSH 370
Qy 387 --SOLQELVSKENSIKDLVNATLARNLRFSLGKYNFLPDDL--ASHLDVYPLVSKAKIKQ 442
Db 371 TKASISNLNKKENIYD-----FGKNGKFNDRLNSPNLEYSLDASASLDK 417
Qy 443 SSITKGLFIELPIKLSKSSILGQ--BPNTK-TLFEKEVTFKLDNFRDVEIEKAFGLLYP 500
Db 418 KD---KSIVLIPYRLIEIKDPFADLLYPDTKDNILVKEGILKLTGFK-----KGPKIDLP 469
Qy 501 GVNELEQARQAQRASTEKEKSKGLKEPQSKTEKSKAINNQGLEBDDNITELPENS 560
Db 470 NINQOI-----FKTEY---LPFFEKGKEQAKL-----DYGNIL----- 500
Qy 561 PIQYQENAGLQASDPKPYMIKDVQYQRYLAKSIOELIKAKDYTKLAKLLNHRHTYNI 620
Db 501 -----NPYNTQ-----LAKVEEALFKGNKQNEIYQALDGNAYAEF 536
Qy 621 SLRLKEQLFDVNPRI--PSSRDIBK-----AKFVLDKTEKNKYWQIYSS--ASP 665
Db 537 G-AFKSVLNSWTGKIQHPKADIQRFTRHLLHGVKIGNSVLNQPOQTAKE-QVISLSKSN 594
Qy 666 VFQNKWSLFGYRYLLGLDPKQTHLVLQKAGL-----QF-EGYENLPSPDNFLDLK 719
Db 595 PFKQGHQVASYFDQLLTQDKLTVLLETLYDLAKKQWLETNQAQFPKGAQYQTKDIFAEADK 654
+Qy 720 NIRIKTPLFSQDNF---KLSLLDENNYDGE--IKAPEGLPLFLPKELRNRSSNGGS 774
Db 655 ---LKFWLKKKKQPNYQINEIHLQSLFNILARNDVIKSDGFGVGVLLLPQSVKTELEGN-- 709
Qy 775 QNSNSPWQEIIISQFKQDNLSNQDLAQFSTKIWEKIIGDENBFDQ--NNRLQYKLLKDLQ 833
Db 710 -----EAQIFEALKKYSLIEN---SAFKTILDKNLEGTDFKTFGDFLKAFFLKAQ 759

```
QY 834 ----ESWINKTRDNLWTY-----LGDKLKVPKKNLEAKFRQISN-LQELL----- 875
Db 760 FNNFAPWA-KLDDNLQVSFEAIKGETTKEGKREBVDKVKELDNKIKGILPQPPAAKPE 818
QY 876 -----TAFYTSALSNWNNYYQDSGAKSTIIFEBIAE---LDPKVKKVCVA----- 918
Db 819 AAKPVAAPAAKPVAA-----KPEAAKPVAAKPEAAKPVAAKPEAAKPEAA 870
QY 919 -----DVYQLKHYAIGFDNAGKFNQEVIRSSRTIYLKTSKGSKLE 961
Db 871 KPVATNTGSLANKPKEDYFPMAFSYKLEYTDE-----NKLSLKTPENLVFLELVHQSEYE 926
QY 962 -ADTIDQLNQAVKNAPLGLQSYLDTERFVGQKLATSLAVQHKQKTEKLPKLNNDGYT 1020
Db 927 DQKIIEKELDKTVNLQYQFQEVKVTSDQ---YQKLSHPMMTEGTQNGQKKGEGTPNQ-- 981
QY 1021 LIHDKLKXVPVTPQISSSEKDFWPEGLNQ---NGSQNVNVSTP-----GSIIESPY 1069
Db 982 -----KKAEGAPSQGKKAEGTPNQKKAEGTPNQSPSTSELTNYPDLGKKKIDELIKKQ- 1035
QY 1070 FSTNFQEDADLDQDG-----QDSDRQGNNSLDNQEAGLLKQKLAILLGNQFIQ 1117
Db 1036 -GKWKTEVELEJEDNIAGDAKLLYFILPDDSKSG-----DPKSSLKVKITVKQSN---- 1085
QY 1118 YYQONDKIE 1127
Db 1086 ----NQBELK 1091

RESULT 24
QYUE9 PRELIMINARY; PRT; 2770 AA.
AC Q7YUE9
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Normocyste-binding protein 1.
GN Name=NBPI;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Rayner J.C., Barnwell J.W., Galinski M.R., Tran T., Huber C.S.,
RA Corredor V.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF533700; AAQ0315.1; -
DR InterPro; IPR011591; Botulinum.
DR ProDom; PD001963; Botulinum; 2.
SQ SEQUENCE 2770 AA; 333744 MW; CC09878E019A86DC CRC64;

Query Match 4.0%; Score 387.5; DB 2; Length 2770;
Best Local Similarity 18.7%; Pred. No. 3.3e-07;
Matches 355; Conservative 325; Mismatches 681; Indels 533; Gaps 89;

QY 50 IDSVAFKPSIANFT---SDYQSVKALLNG-----KTFDPKSEFTDFVSKF 93
Db 779 IDDIKIQETLQITHIVNNIKTIKKDLKKEFIQHLIKYMNERYQNMQGVNLTNYINQY 838
QY 94 DFLTNGTGTVELEIPKQYQVISESPEDDKERFLGHLEKLEDG-----NIAQSATKFI 149
Db 839 EEEENNKKQYITTRINQIKIYDNIYAKEBE-IRSGQYQDFITSRKNYINIRENSKNV 897
QY 150 YLLPLDMPKAALGOVSYIVDKNFNNLIHPLSNFSAQSIKPLALTRSSDFTAKLNQFNQ 209
Db 898 DMIRKEEKK-----KIQNCVDK-YNIS-----KQYVKMLKNGDTQD-----ENNNNN 939
QY 210 DELWVYLEKFPDLBALKANIRLOATDPSFEGKGNLVDPFVYSFIRNPQNKQEWASDLNQDQ 269
Db 940 D---IYDKLIVPLDSIKQNDKYNTENF-----ITFTKINTHNKK-----NQE- 981
QY 270 KTVRLYLRTFSPQAQKTLKDYKYKDETFLS--SIDLKASNGTSILFANE-NDLKDLQ--- 323
```

```
Db 982 -----MMEBFI-----YAYKRLKILKILINISLKACEKNKNSINTLNDKTQELKKI 1026
QY 324 ---DVLDDVSDYFGGSETITS-----NSQVQVPVPAERSLKDVRKFKKDDQOKP 370
Db 1027 VTHEIDLL-----QKDILTQISNKNVLLNDLLKEI---EYIIDIHVHLKAKKSN-- 1073
QY 371 RIKFSIYEDALSFYSQ-----LQELVSKPN---SIKDLVNATLARNLRFSLGKYNF 420
Db 1074 --DLFTYEQSKNVYFKNKKDNFDIOKTIKNQNEWLAIKNYINEINKVQTLYEKKINV 1131
QY 421 LPDDLASHLDYYFLVSKAKIKQSSITKLFILPIKISLSSILGDOEPNIKTLFEKEVT 480
Db 1132 LLNNSKYVQYFY---DHIIINILQKKNYLENTLTKIQ-----DNEHSUYALQQNEEY 1182
QY 481 FKLDNFRD-----VEIEKAFGLLYPGVNEBELQA-----RKAQASPEKE 520
Db 1183 QKVQNEKDQNEIKKQIEKNKNDILTYENNIEQIEQKIELKTNQAKDDQIVNTLNE 1242
QY 521 KSKGLKBFPSQOEKNSKALNN-QEG-LEBDDNITERLPENSPIOYQOEAGLQASDPKP 578
Db 1243 VKKKIITYEKVNDQISNVLKNYEEGKVEYDKNVQNVNDADDTNDIDEINDEINDID 1302
QY 579 YM-----IKOVQORVYLAKSIQIELIKAKDYTKLAKLLSNRHTYNTISLR--LK 625
Db 1303 EINDIDEINDEIKDIDHIKHFDDTYKFDIIVHADD-----TRDEYHIALSNYIK 1353
QY 626 EQLDVNPRISSRDIEKAKFVLDKTEKNKYQIYSSASPVFQNKWSLFGYRYLLGLDP 685
Db 1354 TELRNIN-----LQEIKNIIKIFKPEKSAHKEKESEQINKEFTKMDVVINQLRDI- 1407
QY 686 KQTHLHVKLQKAGLQFEGYENLPSDFN---LEDLKNIR---IKTPLFQKONFKLSLL 739
Db 1408 ROMLDLYKELDEKY-----SEFNKTKIEENINRENINNNVIEWIEKNIIEFLR 1456
QY 740 DFNYYDGETKAPFGLPLF-----LPKELR-----RNSNS----- 771
Db 1457 HMDDQDKAAKYME-NIDTYKNNIEIISKQINPENYVETLNKSNMYSYVEKANDLFYKQI 1515
QY 772 ----GGSQNSNPW-----EQEIIISQFKDQNLSDQDLAQFSTKIWEKIIIGDENEFDQ--- 820
Db 1516 NNIIINSNLKNEAFTIDELQIQKRNKLLTKKQIIQVYTNEL-----ENIFNEIKN 1568
QY 821 -NNRL---QYK-LLKQLOESWINKTRDNLWTYLGDKLVKPKKNLEAKFRQISNLEQLL 875
Db 1569 INNILVTNYKSILODISQ---NINHVSITYEQL-----HNLKILEE--EKSQMK 1614
QY 876 TAFYTSALSNWNNYYQDSGAKSTII-FEBIAELDPKVKKGVADYVQLKPHYAIGFDN 934
Db 1615 TLHKSNVLHQINFNEDAFINLLINIEKIKNDITHIKEK--TNIYIMDVNKS---KNN 1669
QY 935 AGKFNQEVIRSSRTIYLKTSKGSKLEADTIDQLNQAVKNAPLGLQSFYLDTERFVGQK 994
Db 1670 AQLYFHTNLRGNEKIEYLKLNKSTNQIITLQELKQVOENVE-KVKDIYNQTIKY----- 1723
QY 995 LATSLAVQHKQKTEKLPKKNL--NDGYTLIHLKLEKVPVQISSSPKDFEGKLNQNGQ 1052
Db 1724 -----BEKKNYHIITDENKINDILHNSFIKQINME-----SSNNK 1761
QY 1053 SQN-----VNVSTFGSIIESPVSTNFQDAD-----LDQDQDQ---DSRQGNNS 1094
Db 1762 KQTKQIIDIINDKTFEEHIKTSKTNMLKEQSQMKHIDKTLTNEQALKLFVDINSTNN 1821
QY 1095 LDNOEAGLLKQKLAILLGNQFIQYQONDKIEE-FEIIINVEKVEL----- 1139
Db 1822 LDNMLSEINS-----IQNNIHTYIQEANKSFQFKIICDQNVNDDLNLKSLGLDNTMNH 1875
QY 1140 -----SFRVEKL--AKTLEDNGKTIIVLS-----DETMSLIIVNTTIEK 1176
Db 1876 LKNLQNEIRNMNLEKNFMDKSKIDEEBKDLKLVNINNNSLDKLKKYEEALFQK 1935
QY 1177 TPMSAVPEVDFTKWVEQYDPTPLAAKTKFVLKFKDQIPVDGS----- 1220
Db 1936 VKEAEIQENIEKIQEINTLSDVPFKKPPFFI---QLNTDSSQHEKDINNNTVYKNN 1991
```



```

653 KNYTWOIYSSAPVFNQKWSLFGYRYLLGLDPPKQTHLHVLKQKAGLOFEGYENLPD 712
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
869 KKEQSFNDKSLNETKNSIEKEYQNTLKKVDEYIKVKSTKESITKFSKQTIKLD 928
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
713 FNLDELKLR- - - - -IKTPLFSQKDNFKLSLDFNN- - - - -YVDGEI 749
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
929 MLNQNIKTVKETNSIDKSYIEKPEQLTKQTKLENKFTFSLNHEANNNELIKYFSDL 988
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
750 KAPEFGLPLFLPKELRRSNSGGSNSPWEQEIISQFKDQNSDQLAQFSTKIWE 809
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
989 KA- - - - -NLGINEENMLYNQFTB- - - - -KETFNDIKEKNIHNEEISKIEIKIHA 1034
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
810 KIIGDNEBPDNNRLQYKLLKDOESWINKTRDNLTYWYGLDKLVKPKNNLEAKFRQIS 869
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1035 SIYNISEETEREIGIN- - - - -IESLNTKVFVKENV- - - - -TNL-NKIKEKLKHYDFSDFGREG 1088
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
870 NLQELLTAFT- - - - -SAALSNNWNYQDSGAKSTIIPERIAELDPKVKKEV 916
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1089 NIK- - - - -YTDKIKINDIMAVSQIDQHINGLDDIQKSESYYSEMKEQINKL-EKV 1141
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
917 GADVYQLKPHYAIGFDDNAG-KFNQEVIRS- - - - -SSRTIYLK- - - - -TSGSKLEADT- - - - - 964
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1142 S- - - - -NTEISDNVVEGKKKQIIVTKIKKKNIYEEINKLSEISKIEKDNITSLE 1193
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
965 - - - - -IDQLNAQVKNAPLGLOSP- - - - -YLD- - - - -TE- - - - -RP 989
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1194 KVKDINLSYQGNLGNLFLEQIDEEKKAENTIKMSBAYIDDDLNKIKKQSEIETEMDIKM 1253
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
990 GVPQKLATSIAVQHKQKELTKPKLAN-DGYTLIHDKLKKPVIPQISSPEKDWFGKLN 1048
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1254 DI-NKEMEALKISHDDKKCHDKSKNKENISIDYKSSK-IIQDPSRESDDINDIKNKIQ 1311
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1049 QN-GSQNVNVSTFGSIIIESPYSTNFQEDADLDQDQDSRQGNNSLONQEAAGLKQKL 1107
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1312 KVVSESONHN- - - - -SDINQCLNEVANIYILNKNI 1343
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1108 AILLG--NQPIQYQQNDKEIEFIEINVEKV- - - - -SELSFR-VEFKLAKTLEDNG- - - - - 1155
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1344 KKIIVKVEYTSIRKKNKINDELANNSEKVIKIEGDSLKSECRKINSITLDDKDIDEC 1403
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1156 -KTIRVLSDETMSLIVNTT--IEKTPMSAVPVFDTKQVEQDPRTPLAATKTKVLKFK 1212
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1404 IKNIIVLKNILNEETNTHFKNAEYNI-KVLSNPNIEAD- - - - -NKSQYILEIX 1456
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1213 DQIPV-DGSGNISDKMLASPLVIHQOMLRSLPVVKTIRELGLTKBQQOQQOQQOQQO 1271
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1457 KNGTNDHDYNIKE- - - - -LKSHKD- - - - -KSGYKTEADQ- - - - - 1487
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1272 QKKAARKEBEL-ETYNPKDEFNINPLTKAHLTLNLVNDPNYKIEDLVIKNEAGDH 1330
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1488 NKKAIQNKELFEQY--KEBVTVL--LNKYAYAVELKNKFKTKNDKSKQIIEIKD- - - - - 1538
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1331 QLAFSLRANNIKRLMNTPIITFADVNPFPYNNEDWRSIDKVLNKNVSSH-QQOAAAGNQ 1389
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1539 -AHNYCTL- - - - -ESGKS-EKKNKIEKKEIHIEDEVANNDK 1573
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1390 GSGLIQRLNKRIP- - - - -ETFTPALIALKDRNN--TNLSNYSDKLIIMTKPYLVE 1438
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1574 SNKAITSIKVSVPFKTKIINKINEIRTKSDCLKETNDLEKQISNLS- - - - -IDTQETKLTE 1630
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1439 RSGVPSWSTGLDYGISEQTKDGTSSSQKGFQDFIQALGLKNTYHGLSLRIFD 1498
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1631 N--GKQLKT-LEELLESK-KQKNIBDQKDELDEVNSKIKNTENTVNHQKNYEGIGIVE 1686
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1499 PGNELAK- - - - -IK- - - - -DASNNKGB- - - - -EKKL 1519
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1687 KINEIATKNQIBESTKELIKPTIOHIISSFNANDLEGIDSDENLGYNTMGNIVEEFI 1746
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1520 KSYDLFKNYLNEYEKKS- - - - -PKIAGKWTNI- - - - - 1546
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1747 KSYNLITNYLTVSKESIYTIQNIQNIDTQKELLKNIEVNNKAKSYLDYIKENEFDRIV 1806
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

1547 -HPDQK- - - - -EYPNPNQKLPENYLNL- - - - -VLNQPKVTLNSSDFITNLFVPEBGRG 1597
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1807 THFKKLTVDNDFKNEYSKVNEGFDNISNIN- - - - -TVKSTD- - - - -ENSLN 1852
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1598 SGTKLKQVIOQVNNYADWGSAYLTFWYDNIITNQPNVI-TANI- - - - -ADVFIKD 1649
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1853 ILNQTKEMYANIVNNTY- - - - -YSYKYEAHIFRNIKPLANTLIKIKNSSGIDLFP-KD 1905
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1650 VK- - - - -ELENTKLIAPN- - - - -ITQWPNISGS- - - - -KEKFYKPT 1682
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1906 IKIAILSYLDSKTEDTLIPISPOKKTETTYTKISDSYILDLKKSQELQKEQOTLKL 1965
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1683 VFFGNWENENSSMSQAOTPTWEKIREGFALQALKSSFDQKTRTFVLTNAPLPMKYGP 1742
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1966 IF- - - - -ENREL- - - - -YEKQATNELRGTL- - - - -DLKYKKEKILSEVKLLHKSNE 2009
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1743 LGFQNGPNFKTQDWLRFQNDNDNQIAALRVQEQDRPEKSSDKQKWKFKPVIPPEMF 1802
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2010 LN-KLSCNFPQNYDTTILESSKYD- - - - -QVKEKSNNYKQEK- - - - -LGIDFNVTDMEEKF 2059
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1803 NSGNIRFVGVQMOIQCPNTLWLPVINSVIVPYRGTDGSDNDVANLNVPQVKTIAFTNN 1862
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2060 NN- - - - -DVKVIELENNYDSSENNILQSKQKUKEL- - - - -TN 2094
Qy      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1863 AFNNVKEFNISKIIVE 1879
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2095 KFNABIKK--IDDKIIE 2109

RESULT 27
Q7RR9 PRELIMINARY; PRT; 2757 AA.
AC Q7RR9;
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequende update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Rhostry protein.
GN Name=PY00649;
OS Plasmodium yoelii yoelii;
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Anguolui S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.B.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates P.R., III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman D.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter C.J., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: AABL01000178; EAA17609.1; -.
DR InterPro: IPR001623; DnaJ N.
DR InterPro: IPR006499; ReticulocyteBP.
DR TIGRFAM: TIGR01612; 235Kda-fam; 1.
DR PROSITE: PS00636; DnaJ_1; UNKNOWN 1.
SQ SEQUENCE 2757 AA; 324332 MW; 55AD73DE2B9EFA37 CRC64;

Query Match 3.9%; Score 384; DB 2; Length 2757;
Best Local Similarity 18.6%; Pred.No. 4.5e-07;
Matches 410; Conservative 382; Mismatch 725; Indels 686; Gaps 105;
Qy 59 IANFTSDYQSVK- - - - -KALLNGKTFDPKSSSEFTDFVSK-FDFLTNGRT 102
```


[illegible]

Query Match		3.9%; Score 380.5; DB 2; Length 2977;
Best Local Similarity		18.4%; Pred. No. 6.9e-07;
Matches 351; Conservative 329; Mismatches 672; Indels 551; Gaps 89;		
Qy	50	IDSVAFKESIANFT---SDYQSVKALLNG-----KTFDPKSEFTDFVSKP 93
Db	775	IDDHKIQETLQKQITHVNNIKTKIKDCLKKEFIQHILIKYMNERYQNMGOYNNLNYNQY 834
Qy	94	DFTLNNGTVLIEIPKQYQWVISEPDEDDKERFLGPHLKEKLEJEG---NIAOSATKFI 149
Db	835	EEENNMMQYITIRNIQKIYVDNIYAKEKE-IRSGQYVKDFITSRKNIYINIRENISQV 893
Qy	150	YLLPLDMPKALGQVSYIVDKNNFNNLIHPLSNFSAQSIKPLATRSDFIAKLNQFNQ 209
Db	894	DMIKNEEKK---KIQNCVDR-KYNSI-----KQYVKFKNGDQD-----ENNMMNN 935
Qy	210	DELWVYLEKFFDLKALKANIRLOTADFSFEKGNLVDPPVYFIRNPQNKEWASDLNQD- 268
Db	936	D---IYDKLIVPLDSIKONIDKYNTEHNF-----ITFTNKINTHNKK-----NQEM 978
Qy	269	-QKTVRLYLRTEFSQAATILKDYKYKDETFLSSIDLKASNGTSLFANE-NDLKDQL--- 323
Db	979	MEEFIVYVKLKL-----ILK-----ILNLSLKACEKNKSINTLNDKTQELKKI 1022
Qy	324	---DVDLLDVSDYFGGQSETITS-----NSQKVPVAPASRSLKDRVKFKQOQKP 370
Db	1023	VTHEIDLL-----QKDILTQISNKNVLLNDLLKEI---EQYIIDVHKLKKGNN-- 1069
Qy	371	RIEKFSLEYDALSPVSO-----LOELVSKPN---SIKDLVNATLARNLRFSLGKYNF 420
Db	1070	---DLFTYEQSKNYPYFKNKNDFDIQKTINKMNEWLAIKNYINEINKYQTLVEKKINV 1127
Qy	421	LPDDLASHLDYVFLVSKAKIQSSITTKLFTLEPIKISLKSILGDOBPNIKTIFXEVT 480
Db	1128	LLHNSKSVQVIFY---DHIINLILQKNYLENTLTKIQ-----DNEHSLVALQONEEY 1178
Qy	481	FKLONFRD-----VETEKAPGLLYPGVNELEQA-----RKAQASFEKE 520
Db	1179	QKVNEKDQNBKIKQIKIENKNDILTYENNIEQIEQNKIELTNAQNKDDQIVNTLNE 1238
Qy	521	KSXKGLKFEQSKENSKAINN-OBG-LEEDNDITERLPENSPIQYQOENAGLGASDPKP 578
Db	1239	VKKLIITYEKVDNQISNVLKQYBEGKVEYDKNVQVQNVNDADDNDDEINDIDEIND 1298
Qy	579	YM-----IKDVQQRYLAKSQIOBLIKAKQDYTKLAKLSNRHTYNISLR--LK 625
Db	1299	EINDIDEINDIDEIKDIDHIKHFDOTKHFDIYHADD-----TRDEVHIALSNVYK 1349
Qy	626	EQLFDVNR-----IPSSRDIEKAKFVLDTKEK--NKYW-----QI 659
Db	1350	TELNRINLQEIKNNIKIITKEFESAHKEIKESQINKKEFTKMDVINVQLDRIDRQMLDL 1409
Qy	660	YSSASVPQNKWSLFGYVRYLLGLDPKQTIHELKVLGQKAGLQFEGYENLPSDFENLEDLK 719
Db	1410	YKE-----LDEKISEN-----KTKIEINNIRENNINNVIEYKNIIEYFLRHMN 1455
Qy	720	NIRIKTPLFSQK-DNFK-----LSLLDFNNYYDGEIEKAPFGLPLFLPKE 763
Db	1456	DQKOKAAKYMENIDTYKNNBIEIISKQINPENYVETLNKSNMYSYVERKAND---LFY-KQ 1510
Qy	764	LRRNSSNGGQNSQNSPW---EQELISQFKDQNSQNDQLAQFSTKWEKIIGDENEFQD 820
Db	1511	INNIINS--NQLKNEAFTIDELQNIQKRNKLLTKKQOIIQYTNIEI-----ENIFNE 1561
Qy	821	---NNRL---QYK-LLKDOESMINKTRDLNLYTLGDKLVKVPKNNLEAKFRQISNLQ 872
Db	1562	IKNNILVLTNYKSILODISQ---NINHVSIYTEQL-----HNLVYKLEB--EKE 1607
Qy	873	ELLTAFYTSALSNMNNYYQDSGAKSTII-PEEIAELDPKVKYKGVADVYOLKPHYAIGF 931
Db	1608	QMKTLYHKSNSVLHQNINFNEDAFINLLNINTEKINDITHIKEK--TNIVMIDVNSK--- 1662

Qy	932	DDNAGKFNQEVIRSSRTIYILKTSGSKLEABTDIDQLNQAVKQAPLGLQSFYLDTERGV 991
Db	1663	KNNAQLYFHNLTARGNEKIEYLNKLNKSNSTNQOITLQELKQVQENVE-KVKDIYNYQTIKY-- 1719
Qy	992	FQKLATSLAVQHQKQKXTLPKCLN--NDGYTLTHDKLKPKVPIDPOISSSPKDWFEGLNQ 1049
Db	1720	-----BEKIKNKHIIITDENKINDILHNSFIKQINWE-----SS 1754
Qy	1050	NGOSQN-----VNVSTFGSIIESPYSTNFQEDAD-----LDQDGD-----DSRQG 1091
Db	1755	NNKQTKQIIDIINDKTFEEHNTSKTKINMLKEQSKMKHDKTLLNEQALKLPVDINST 1814
Qy	1092	NNSLDNOEAGLLKQKLAILLGNQFIQYQONDKEIE-PEIINVEKVEL----- 1139
Db	1815	NNNLDNMLSEINS-----IQNNIHTYIOBANKSFDKFKIICDQNVNDDLNLKLSGLDLY 1868
Qy	1140	-----SPRVEFKL--AKTLEDNGKTIIRVLSDETWSLIVNTTIEKTPENSAV 1183
Db	1869	MNHLNKLQNEIRNMNLEKNFMDKSKIDEEKKLDILKVNISN--INNSLDKLKKY--Y 1924
Qy	1184	PEVDTKWBQYDPRTPLAATKFPVLK-----FKDQIPVDGS----- 1220
Db	1925	EEALFQVKVEKAENQKENIEKIQEINTLSDVFKKPPFFIQLNTDSSOHEKDIINNVEY 1984
Qy	1221	-GNISDKWLASIPLVHQOMRLSPV--KTIRELGLKTEQOQOQOQOQOQOQPKKA-- 1275
Db	1985	KNNIDE--IYNVFIQSNLIQKYSSEIFSTLANYIQTKEIKESIKEQNLQNLQNEKEAFV 2042
Qy	1276	-----VRKEELETYNPKDBFNILNPLTKAHLRLTSLN-----VNNDP 1313
Db	2043	LLNKKINETIKFLQKQIKNERQNDVHNIKEDYNLLQYLYNMKNMEMQLKKNYNDVHMWK 2102
Qy	1314	NYKIEDLVKINBAGDHQ-----LAFSLRANNIKELMNTPTITFADYNPPFYNNEDMRS 1366
Db	2103	NY-----VENNNGEKEKLLKETISSYDYKINNINNKL-----VIYKKE-- 2141
Qy	1367	IDKVLANKGNVS-----SHQOQAAAGNGSGSLIQRNLNKNIKPETFTPALIALK 1414
Db	2142	-DYFNNMIKVSILNIIHKKQOQNEQIRVIVNABYDSSLINK-DEEIKKE-INNOIIEI- 2197
Qy	1415	DRNNTLSNYSDKIIMIKPKYLVERSIGVPWSTGLDGYIGSEQTKDGTSSSSQKGFDDQ 1474
Db	2198	NKENENISNFKDQIQNKQ----- 2220
Qy	1475	FIQALGLKNTYEGKGLSIRIFDPGNE-----LAKIKDASNKKE--EKLK----- 1520
Db	2221	IITNM-----NDMEKSTILLVDIIQKKEEALNKQKNILNRNIDNILNKKENIIDKVIKNC 2276
Qy	1521	-----SYDLFKVNLNEYEKSPKIAKGWNTNHPDQKEYPNPNQKLPENY----- 1564
Db	2277	DYKDILIONETEYQKQONINHTYEEKKSI-----DILKIKNIKQKNIOBYKNKLE 2327
Qy	1565	-LNLVLNQPKVTLNYSNDFITNLFEVEPGSBRGSGTKLKQVIO-----KQVNNNYADW 1617
Db	2328	QMTIINQSIQEHQVFINADILQ-----EKIKLEIILKNLDILDEQIMTYHNSI 2376
Qy	1618	GSAY-LTFYVDKNIITNOPVITANIVFIKQVKELEDNTKL 1659
Db	2377	DELYKLGIOCDNHLITTSVNVNKNITKIMIHKKQKQEDIQKI 2419
RESULT 30		
ID	Q9BJX9	PRELIMINARY; PRT; 2747 AA.
AC	Q9BJX9;	
DT	01-JUN-2001	(T-EMBLrel. 17, Created)
DT	01-JUN-2001	(T-EMBLrel. 17, Last sequence update)
DT	01-MAR-2003	(T-EMBLrel. 23, Last annotation update)
DE	235	Kda rhotry protein (Fragment).
GN	Name=Iia.2;	
OS	Plasmodium yoelii yoelii.	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX	NCBI_TaxID=73239;	

Qy	1684	FFGNWENENSNMSOQPTTWKREGFALQALKSSFDQKTRTTFVLTTNAEPLPMWKYGPL	1743
Db	1968	F-----ENRRL-----YKQATNELRGTLSS--DLKYKKEKILSEVKLLHKSNE	2011
Qy	1744	GFQNGPNFKTQDWRLVFQNDNDNQIAALRVQODRPEKSSDKQKWKFKPVVPEEMFN	1803
Db	2012	N-KLSCNFQNYDTILESKEYD-----QVKEKSNYYQBEKE-LGIDFNVDMEEFKN	2061
Qy	1804	S 1804	
Db	2062	N 2062	
RESULT 31			
Qy	Q6YA79	PRELIMINARY; PRT; 3427 AA.	
AC	Q6YA79		
DT	05-JUL-2004	(TREMBLrel. 27, Created)	
DT	05-JUL-2004	(TREMBLrel. 27, Last sequence update)	
DT	05-JUL-2004	(TREMBLrel. 27, Last annotation update)	
DE		Reticulocyte binding-like protein 2b.	
GN	Name=RBL2b;		
OS	Plasmodium reichenowi.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5854;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	PubMed=14698440; DOI=10.1016/j.molbiopara.2003.10.017;		
RA	Rayner J.C., Huber C.S., Galinski M.R., Barnwell J.W.;		
RT	"Rapid evolution of an erythrocyte invasion gene family: the		
RT	Plasmodium reichenowi Reticulocyte Binding Like (RBL) genes.";		
RL	Mol. Biochem. Parasitol. 133:287-296 (2004).		
DR	EMBL; AY156677; AAC38039.1; -		
SQ	SEQUENCE 3427 AA; 404039 MW; F7C66ED1296686C1 CRC64;		
Query Match 3.9%; Score 378.5; DB 2; Length 3427;			
Best Local Similarity 19.8%; Pred. No. 1e-06;			
Matches 412; Conservative 307; Mismatches 750; Indels 609; Gaps 101;			
Qy	66	YQSVKALLNGKTFDPKSSSEFTFVSKPFDLTNNGRVTLEIPKYYQVVISFSPEDDK--	123
Db	872	YENLDKK-LKQNTV-PNEWHHTSLVQKTEYITVEDKTISELEQ-----EFNNNEKLD	922
Qy	124	-----ERFLGFH-----LKEKLEDGNIQAQATKFI-----	149
Db	923	NILQDINAMNLINILQTLNTGINCRNTNNKNVHLLNKKIELNLNDQMIKNDII	982
Qy	150	-----YLLPLDMPKALGOYSYIVDKNFNL--IIHPLSNFSAQIKPLALTRSSDFI	200
Db	983	QDNKEKNFLAVLKKEEKLEK--ELDDIKLNNLKNWIKHLNLSYDSKQNI---ESNDKI	1037
Qy	201	AKLNQFNQDELWLYLEKFFDLEALKANIRLQADFSF--EKNGLDPFVYVSFIRNPQN	258
Db	1038	NLDSFEKEKDSWVHFKSDIDSLVVEYNICNQTHTNTIKQKNDIIE-LIVKRIK----	1090
Qy	259	KEWASDLNOD--QKTVRLYLRTESPQAKTILKY-----KYKDTFFLSDDLKASNG	309
Db	1091	-----DINQEIITKKVDNYY---SLSDKALTKLSIHFNIDKEKYNPKPSQENIKLEN	1142
Qy	310	TSLPANENDLAKQI-----DVLLDDVSDYFGQSET-----	340
Db	1143	MILEKKINEDKALIQIKNLSHDFVKADNEIKKEEVD--DQTHYSEKRVGMGIYK	1200
Qy	341	-----ITSNSQVXPVPASERSLKDRV-----KFKKQDQOKPRIEFK--	375
Db	1201	DIKKKLDELANNKLIDITLNEANKIESEYERILLDDICEQITNANKNSDTIKEQIESYK	1360
Qy	376	-----SLVEYDALSF-----YQLOELYSKPNISIK	400
Db	1261	HIDYVDADVSKTSNDHHLNGNIHDSFFYEDTLNKKAYFKLKDYEINIKLTNESGLK	1320
Qy	401	D-----LVNATLARNLRFSLGKYNFLFDLASHLDYFLVLSAKIKQ-----	442

Db 2225 DISIFNKY-----DDYIKVDYMSNNIDVV-----NKHNSLLSEHVINATNIE 2268
Qy 1404 EFTFPALIAKDRNNTNLSYGDKIMI-----KPKYLVERISGVPMSTGLDGYIGSBOT 1458
Db 2269 NIWTCIVQINKDTEMNFTBETKDNLLKYENFKKEKNIINNNYKIVHFNKJKEIENSLET 2328
Qy 1459 KDGTSOSSQKGFQDQFIQALGLKNTYHGKGLSIRIFDPCNELA-----1504
Db 2329 YNSISTNFKISEAQN-----IGLKNVEFN-----NITKINDKVKELVHVVDSTUTLESIQTF 2381
Qy 1505-----KIKDASNNKGSEKLLKSY-----DLFKNYLNEYEKSKSP 1538
Db 2382 NKLYGDVMSNIQDFYKEDINN--GELKKVKLYIENITNLLGRINTFIKELNKYQDENV 2439
Qy 1539 IAKGWTNIHDDQKEY-----PNPNQKLPENVLVLNQPWKV--LYNSS--DFITNLFVE 1590
Db 2440 IDK-YTEINKNNSYIIKLKERANNLKQFNLLQNIKENETELYNINSIKODIMNMWKS 2498
Qy 1591 PEGSDR--GSGTKLKQVI-----OKQVNNYADMGSAYLTFWYDQNIIT 1632
Db 2499 VNNIKQFRNULPKEKLFQMEKOKLLDINNINNETKRISNTDAYNITQDIENNKNKEN 2558
Qy 1633 NOPNVITANIADVIFIKVKELENDNTKLIAPNITQWPNISGSKEKFPKTPVFFGNWENEN 1692
Db 2559 NNNNKT--IDKLIDDIK--IHNEKIQAEILI-----IDDAKRVNEIT-----DNIN 2602
Qy 1693 SSMNSQAQPTTWBKIREGFALQALKSSFDQKTRTFVLTNAPLPLWKYGLFGQNGPNPK 1752
Db 2603 KAPTEITETNNND--NNGVKSAKNIVDEAT--YLNNELOKFLKLKLNELLSHNNNDIK 2656
Qy 1753 -TQDWRLVQNDNDQIAALRVQDQREKSSSEDKQK 1789
Db 2657 DLDDEKLIILKEBERQEKLAQBEERKERERIEKEKOE 2694

RESULT 32
Q26223 PRELIMINARY; . PRT; 2269 AA.
AC Q26223;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Rhostry protein.
OS Plasmodium yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5861;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YM.
RX MEDLINE=95021522; PubMed=7935623; DOI=10.1016/0166-6851(94)90125-2;
RA Keen J., Sinha K., Brown K., Holder A.;
RT "A gene coding for a high molecular mass rhostry protein of Plasmodium yoelii".
RL Mol. Biochem. Parasitol. 65:171-177 (1994).
DR EMBL; L27838; AAA21304.1; .
DR PIR; T28677; T28677.
DR HSSP; P02652; 116L.
DR InterPro; IPR006499; ReticulocyteBP.
DR TrEMBL; TIGR01612; 235kDa-fam; 1.
SQ SEQUENCE 2269 AA; 265158 MW; F3D8CB103FB9A6E1 CRC64;

Query Match 3.9%; Score 377; DB 2; Length 2269;
Best Local Similarity 18.5%; Pred. No. 6.7e-07;
Matches 418; Conservative 363; Mismatches 762; Indels 712; Gaps 106;
Qy 19 STVFGTVGLASKVKYRG-----VNPTQGVISQLGLDSVAFKPSIANFTSDYQSVKAL 73
Db 165 NTIYNTIKSYFDQI-YEGDDIDTFYNELSSIVKE-DPIDDIEDTKLENLSKIDNVYDKI 222
Qy 74 LNKGTDPKSESEFTDFSVKDFDLTNNGR---TWLEIPKQYQVVISSEPSPE-----120
Db 223 -----QKWEIETVKSHLNNIETNNKLPNTILEI-KKY--IYDEISKELNKKMLDFKN 271

Qy 121 -----DDKERFRLGPHLKEKLEDGHIASATKFIYLLPLDMPKAALGOYSYIVD 169
Db 272 KEKELSNKISDYDKKREQLSEYKSKMLEIRHNSQTN-----VDNTKEEAKQNY--D 323
Qy 170 KNFNLLIIHPLSNFSAQSIKPLALTRSSDFIAKLN--OFNNQDELWYLE--KPFDL-E 223
Db 324 KSNEHMTIPTNEDEISKIISGVKTMKDBEILSKVNTYIDFNKKYKETYVNSHQSQTLELTD 383
Qy 224 ALKANI-----RLQATDFSEKGNLVDPPVSYSTRPNQ--QKEWASDLADODOKTVLYLR- 277
Db 384 KIKAEVSDKELKCKEQSFNDK-----SLINETKNIEKEY-QNIN-TLKKVDEYIKV 434
Qy 278 -----TEFSPAQKTIKLD-----YKYKDETFLSSIDLKASNGTS 311
Db 435 CKSTKESITKSSK-OTILKQMLNQNIKTVKETNSIDKSYIEKFPQILTGKQTKLENKFT 493
Qy 312 LP-----ANENDLKQDLDVLLDVSDFGGQSETITSNQVQKVPASERSLKDVRKFKK 365
Db 494 EPLSNHNEANNELTKYFS--DLKANLGINENMLYNQFTE---KEKTFND-IKEKN 544
Qy 366 DOOKPRIEFSLYEYDALSFYSQLOELVSKNSIKDLVATLARNLRSLGKYVFLFDDL 425
Db 545 IHINBEISKIEIKIH--ASIYINISEETREBIGINIESLNTKVFKEKGVNTVNLNKEKL 602
Qy 426 ASHLDYVFLVSKAKIKQSSITKCL---PIELPIKISLSSIIIGDQEPNIKTLFE--KEYT 480
Db 603 -KHVDFSGFEGEKNIKYTDKIKINDDIMASQQIDQHINGLDDIOKKSSEYVSEMKEQI 661
Qy 481 FKLDNPRDVEI--EKAFLG-----LYPGVNEELFOARKAQAASPEKES 522
Db 662 NKLEKVSNTESINDNVEGKKKQIIVTKIDKKNIYBEINKLSEIKSEKDNLSLEKV 721
Qy 523 KK-----GLKEFSQOKEENKAINNOGLEED-DNITERLPE-----NSPI 562
Db 722 KDINLSYQONLNFLEQIDEEKKAENTIKSMEAYIDDLNIDKKKSQIEETEMDKMDI 781
Qy 563 QYQENAGLGSADPKPYMIKVQVQNR-----YLLAKSQIQELIKADY---TKLAKLS 613
Db 782 NKEMALKISHDDDKKCHDKSKNHKENISDIYDKSKIIQDFSRESDDINDIKKLQKNVS 841
Qy 614 NRHTYINSLRLKEQLFDVNPRIIPSSRDIEKAKFVLDKTEKNKQWYIYSSASPVQKNWSL 673
Db 842 ESQNHNSDI--NOCLNEVANIYNILKLNKIKKIIDKVE-----YTSIEKN--- 886
Qy 674 FGYYRYLLGLDPKQTHLKVGLQKAGLPAGVENLPS-----DFNLED-----LKNIR 722
Db 887 -----KKNINDELNNSEKVIKIBGDLSEKRSKINSTLDDDKDIDECIKLIN 934
Qy 723 I-KTPLFSOKNF-----KLSLIDPNN-----YVDGEIKA-----PEFGL 756
Db 935 VLKKNILNEETNITNHFKNABEYKIVLSNPANNIEMADNKSQYILEIKKNGTNDHDYNI 994
Qy 757 PLFLPKELRNSNSGG-----SONSNS--PMBQETISQFDQNLSNQDQLAOFSTIKWK 810
Db 995 -----KELKSHKDKSGYKTEADQNKKAIQKNKELPEQYKEE-----VTVLLNK 1038
Qy 811 IIGDE--NEFDQNNRLQYKLLKDLQESW-----INKT 840
Db 1039 YVAVELKKNKFDTKNDKSKQIIEIKDAHNYCTLESKSEKKNWEIKNEKIHIEDEVAND 1098
Qy 841 RNLVWYTLGDKLKVKP-----KNNLEAKPROISNL-----871
Db 1099 KSNKAITSI--KVSVEPPTKIIKINEITRSDCLKETNDLE---KQISLNSIDTQETK 1153
Qy 872 -----QELTAFYTSAAALNNNNYVODSGAKSTIIFERIAELDPKVEKVGADVYQKPH 926
Db 1154 LTENGKQLKTEBELLESKKQKNIEDQK-----KELDEVNSKITK-NIENTVNOHKKN 1205
Qy 927 YAIIGF-----DDNAGKFN---939
Db 1206 YEIGIVEKINEIAKTNNKQIESTKELIKPTIOHIISSFNANDLEGDSDENLKGKNTYEMG 1265
Qy 940 ---QEVIRSSSR-TIVLKTSGK-----SKLEADTIDQLNQAVKNAP-LGLQSFYLDTERF 989

[illegible]

Db 1061 INKLDEISE-----IEKNNTSLKKVKDINVYQSGLGKLFLE---QIDBEKKKAGHM 1110
 QY 468 EBNITKPEKVTFLDNFR-----DVEIEKAFGL---LYPGVNEELE----- 507
 Db 1111 ---IKSM---EYMDLNLNKKQSIEENEMKINMDIKMY--INKEMEALNISHDDDKQVH 1163
 QY 508 --QAKAQAASPEKSKGLKEFS-----QKENSXAINN----- 542
 Db 1164 NISKKHEEKISDIHKNLSKIIQEFSTESINDIKNELQKVSERKNYNNNDINQIILSKIAS 1223
 QY 543 -----QEGLEEDDNIITELPENSPYOEN-----AGLGA 573
 Db 1224 IYNILKLNKIKIIDVKVEYTHEIEKNKNITVELNSKTIIEKEEDSNLKDCKSKIGS 1283
 QY 574 SPDKPYM---IKDVQNRQRYLAKSQTOELIKARDYTKLAKLSNRHTYINISLRKEQLPD 630
 Db 1284 TIDDDVVGECIKITDLTKYILNKE-----TNIAHFNAEKYENVLL----- 1327
 QY 631 VNPRIPSSRDIE-----KAKFVLDKTEKNKYQWYSSASPVFQNKWSLFGYRYLLGLDPK 686
 Db 1328 -----NPNNIEMADNKAQYIL-KIKKN-----NGTNHHY----- 1356
 QY 687 QTHIELVLGQKAGLQFEGVENLPSPLENDLKNIRIKTLPFSOKDNFKLSLLDFNNYD 746
 Db 1357 -NITELKKHKOMS-----NGYKN-EADTNIQAFKK--NKELEKYKQDVTVLDD--KYV- 1404
 QY 747 GEIKAPEFGLPLPKELRRNSNSGSGNSPWEQEIISOQKQN-----LSNQDLAQF 803
 Db 1405 -----AJALKNKFKDKNDS-----EQIIEIKQHTHTTFISOADKSEQ 1443
 QY 804 STKIWEKIGDENEFQNNR-----LOYKLLKOLQESWINKTRD-- 842
 Db 1444 MKEIKNEKINIEDEVAKNDSKNAIIDIQLSVQPFETKFLKIDLTYSDDCLKETKTE 1503
 QY 843 ----NLYWTYLGDKLVKPKNLEAKFROISLNOELLTAPYTSAAALSNNWNYQDSGAKS 898
 Db 1504 KEISNLSI---DTQETKLENGE---KDLTEKLI-----ESLKNKKYIE--GKK- 1546
 QY 899 TIIFEBIAELDPKVEKGVADYVQLKFHYAIGF----- 931
 Db 1547 ----KELDEVNSKI-EKIKIDVKNHKNVEIGIVEKINEIAKANKQIBESTQKLIKPTIE 1601
 QY 932 -----DDNAGKNQEV-----IRSSRTI-YLKTSGSKLEADTIDQ 967
 Db 1602 KLTSFPTNLDKINTDENLKGNTKNTGMYNEFMSYNRIADYLTKTVSKPEITYDDI-- 1659
 QY 968 LNAQVKNAPLGLOSYLDTERFGVFOKLATSLAVQHKQKE-----KTLPKK 1013
 Db 1660 ----KNTRINAQ-----ELSKIKIENKAKSYLNDIEANBFDRIVTHFKNK 1702
 QY 1014 LN--NDGYTLIHDKJJKPV-----IPQISSPEKDWFEKLNQNGSQSNVNVSTFGSII 1065
 Db 1703 LNTVNDNFKNEYKVEGPNDSINSINNVKNSTDENLLNLNQTQO-----IVANIV 1755
 QY 1066 ESYFTSNFOEDADLDQDQDQSDRQGNLSLDQEAQKLLKQKLAILLGNQFIQYVQNDKE 1125
 Db 1756 NQYYSYKYBAE-----NIPIN-----IPKLANLNIQI-----KNSSG 1789
 QY 1126 IE-FEILNEKVSLSFRVEFKL--AKTLEDNGKTIRVLSDETSMLIVNTIEKTPMSA 1182
 Db 1790 IDLFPKNINAILPHLSQKEDTLTFSPSPHNSIKTYTKISDSYNTLL--DILKRSQELQK 1847
 QY 1183 VPE-----VPTDKVQEYDPRPTPLAAKTFVLKFKQIQIPVDGSGNISDRKLASIPLVHQ 1237
 Db 1848 KEQKTLNLIIFENRL--YD-----KQVATNELKOTL--SNLKNKKEQILNEVKLLLHK 1896
 QY 1238 --OMLRLS-----PVVKTIRELGLKTEQOQQOQQOQQO-----QOQPKAVRKE 1279
 Db 1897 SNEINLKLSCNFQNYDALEPAKYDKIKEKSNYEQEKNKLGIDFDVTAMEKFNNDIKVI 1956
 QY 1280 EELE-----TYPKDB-ENLNPILTAKHRLTSLNVLNNDPNKYIEDOLKVKNEAGD 1329
 Db 1957 EELENNYSSEKNTYNFSENNILQSKQKLKELT--NEFNTEIK-RIEDKVEIKN--GL 2011

QY 1330 HOLAFSLRAN-----NIKRLMNTPTTFADYNPFF-----YYNEDWPSIDKYLNNKGN- 1376
 Db 2012 INKLIEIRKNCMLPHTHTLSETLKIT--DYSEFIASATKFSKFLKYIDDINSLSLND 2069
 QY 1377 -----VSSHQOO-----AAGNQSGSIGTORLNKKNIKPETFPTPALIALKDRNTNL- 1421
 Db 2070 IGTLOIKYDLHQTKYVTSIFAEATNDNDLJE-----KEKEATQAI-----KULTNLF 2118
 QY 1422 ---SN-----YSDKIIMIKPKYLVERSIGVPWSTGLDYGISGQTKDGTGSSSSQKGF 1471
 Db 2119 TIDNSIDANLHNNKLQW-----YFNSSELHKSISIKKLYKYM 2158
 QY 1472 DQDFQALGLKNTYHGLGLSIRIFDPGNELAKIDKASNKKGEKLLKSYDLFKNYLNE 1531
 Db 2159 HVFKLLNLTGQINERY-----FD-----ISKQFDDILOQESSELKENLND 2197
 QY 1532 YEKSPKIAKWTNIHPDQKEYPNPN-OQLPENY-----LNLVLNQPWKVTL 1577
 Db 2198 LKTIQKLSIDKKNKFLNELSETPIPNTLKEIYHIVKYESQIDEIKNITNDENDNITL 2257
 QY 1578 YNSDPFITNLFPVPEPSDRSGTKLKQVIOQVNNNNYADMGSAYLTFWYDKNIIT--NQ 1634
 Db 2258 Y--MDIITKLM-----KKVESIL-----NFVTYEND 2282
 QY 1635 PNVTITANTADVPIDKVELEONTKLIAPNITQWPNISGSEKFKYKPTVFFGNWENENSS 1694
 Db 2283 SNVIKQIHTODNNDVSKIKNLKTIESFKILNKINEIKAQFY-----DNNN 2331
 QY 1695 MNSQAQTPTEKIR-----EGFALQALKSSFDQKTRTFVLTNAPLPLWKYGL 1743
 Db 2332 INNVLSTISQDVIDVKKHISKDLTIENELIBIQKLSLEDIKNSTYDIRSE---QIAKY--- 2385
 QY 1744 GFQNGPNKPTQDWRLVFQNDNQIAALRVQODRPEKSESDKQKWKFKVVIPEMPN 1803
 Db 2386 -VNTIRNVQEQTKIQQNPNKDEIDDIQIQIVNYNKESETK-----LPTITGN 2433
 QY 1804 SQNIRFVGWMOIQGPNLWLPVINSSVYDFVRGFGSDNDVANLANVAPQVKTIAFTNNA 1863
 Db 2434 KNN-----DTSIISRINKFI--NLIKSEYNNNDNVSYNVAK---KLEEDVNI 2476
 QY 1864 FNNVPEFKNISKVIVE 1879
 Db 2477 ILDLKSNQNLKDLIQ 2492

RESULT 34
 QYREV3 PRELIMINARY; PRT; 2749 AA.
 AC QYREV3;
 DT 01-MAR-2004 (TREMBLrel. 26, Created)
 DT 01-MAR-2004 (TREMBLrel. 26, Last sequende update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE 235 kDa rhostry protein (Fragment).
 GN Name=PY04930;
 OS Plasmodium yoelii yoelii;
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=73239;
 RN [1]
 RC STRAIN=17XNL;
 RC PubMed=1236865; DOI=10.1038/nature01099;
 RA Carlson J.M., Anguolui S.V., Suh B.B., Kooij T.W., Perteu M.,
 RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
 RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V.,
 RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
 RA Florens L., Yates P.R., Raine J.D., Sinden R.E., Harris M.A.,
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 RA Carucci D.J.;
 RT "Genome sequence and comparative analysis of the model rodent malaria


```
Db 2257 SIDSIKOLYKQOMHVKLLNIGHINKY- - - - - FDISKEFDNILOQSE- - - - - 2300
QY 1520 KSYDLPKNYLNEYEKSPKIAKGWTHIHDPQKEYPNP- - - - - QKLPENY- - - - - L 1565
Db 2301 - - - - - LKENLNDLKKIGQKISDKKKFLNELSEIPIPNFTNLKEIYVHEIVKYESQIDEIK 2355
QY 1566 NLVLNQPWKVTLNNSDPFTNLVPEBEGSDRGSGTKLKQVIOKQVNNYADWGSAYLTFTW 1625
Db 2356 NITDENDENITLY- - - - - MDIITKLM- - - - - KVESIL- - - - - 2383
QY 1626 YDKNIIT- - - - - NQPNVITANITADVFIKDVKELEDNTKLIAPNITQWMPNISGSKEKYKPT 1682
Db 2384 - - - - - NFTVYENDSNVIKQHLQDNNENDSVIKONLKTIESFKILINKLINEIKAQFY- - - - - 2437
QY 1683 VFFGNMENENSSMNSQAQPTWBEKI- - - - - REGFALQ- - - - - ALKSSPDQKT 1724
Db 2438 - - - - - DNNNNNNVISTISQDVIDVKKHISKDLTIENELIBIQKSLEYIKKSTYDIRS 2489
QY 1725 RTFVLITNAPLPLWKYGLGFGONGPNKPTQDWLVLVFQNDN- - - - - QIAALRVQEQDRPEKSS 1782
Db 2490 EQITKYVN- - - - - PIHDY- - - - - VEQQTKKIQNDPKNDEIDL-IOETVWYNKES 2533
QY 1783 EDKQKQWIKFKVPIPEEMNSGNIRFVGVMQIQGPNLWLPVINS- - - - - SVIYDVPYGTGDS 1841
Db 2534 ELK- - - - - LPTIINNKDNV- - - - - TPIISRIDKVINLIKSEYNN 2567
QY 1842 NDVANLNVA 1850
Db 2568 NDNVSYNVA 2576

RESULT 35
Q8WR56 PRELIMINARY; PRT; 2957 AA.
AC Q8WR56
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Normocyte-binding protein 1.
GN Names=NBPI;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5633;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D42;
RX MEDLINE=21590493; PubMed=11733572;
RA Rayner J.C., Vargas-Serrato E., Huber C.S., Galinski M.R.,
RA Barnwell J.W.;
RT "A Plasmodium falciparum homologue of Plasmodium vivax reticulocyte
RT binding protein (PvRBP1) defines a trypsin-resistant erythrocyte
RT invasion pathway.";
RL J. Exp. Med. 194:1571-1581 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D42;
RX Rayner J.C., Barnwell J.W., Galinski M.R., Tran T., Huber C.S.,
RA Corredor V.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF411931; AAL38220.2; -;
DR ProDom; PD001963; Botulinum; 2.
SQ SEQUENCE 2957 AA; 356102 MW; 3BE03A234E78F52E CRC64;

Query Match 3.9%; Score 376.5; DB 2; Length 2957;
Best Local Similarity 18.6%; Pred. No. 1e-06;
Matches 352; Conservative 321; Mismatches 680; Indels 543; Gaps 88;

QY 50 IDSVAFKPSIANFT- - - - - SDYQSVKCALLNG- - - - - KTFDPKSEFTDFVSKF 93
Db 771 IDDIKIQTLLKQITHIVNNIKTKKLLKEFIQHLIKYMNERYQNMQQGYNNLTNYINQY 830
QY 94 DFLTNNGRTVLEIPKTKQVVISPEPDDKERFLGHLKLELDG- - - - - NIAQSATKFI 149
```

```
Db 831 EEENNMMQYITIRNIQKIYDNIYAKEKB-IRSGQYKDFITSRKNIYINIRENISKNV 889
QY 150 YLLPLDPMKAALGOYSYIVDKNFNNLIHPLSNFSAQSIKPIALTRSSDFTAKLQNFNQ 209
Db 890 DMIKNEKK- - - - - KIONCVDK-YNIS- - - - - KQVVKMFKNGDTQD- - - - - ENNNNNN 931
QY 210 DELWVYLEKFPDLKALKANIRLQADPFSEKGNLVDVPFVYSPIRNPQOKEWASDLNQD- 268
Db 932 D- - - - - IYDKLIIVPLDSIKQINIDKYNTHENF- - - - - ITFTNKIWNTHKK- - - - - NQEM 974
QY 269 -OKTVRLRLTEFSPQAKTILKDYKQDETFLSSIDLKASNGTSLFANE-NDLKDQL- - - - - 323
Db 975 MEPIVYVYKELK- - - - - ILK- - - - - ILNISLAKACEKNKNSINTLNDKTQELKKI 1018
QY 324 - - - - - DVDLDVDSYFGQSETITS- - - - - NSQVFPVPASERSLKDVRVFKDQOKP 370
Db 1019 VTHEIDLL- - - - - QKDILTQISNKNVLLNLLDLEI- - - - - EQVIIDIHLKKKS- - - - - 1065
QY 371 RIEKESLEYDALSFVSO- - - - - LOELVSKPN- - - - - SIKDLVNATLARNLFLSLGKYNF 420
Db 1066 - - - - - DLFTYEQSKNYFYKKNKDNFDIQTINKONWELAKNYINEINKYQTYLYEKKINV 1123
QY 421 LFDDLASHLDYVFLSVKAKIKOSSITTKLFIELPIKISLKSSILGDQEPNITLFEKYT 480
Db 1124 LLHNSKSYQVIFY- - - - - DHIINLILOKKNYLENTLTKIQ- - - - - DNEHSLVALQONEEV 1174
QY 481 FKLDNFRD- - - - - VEIEKAFGLLYPGVNEBEQA- - - - - RKQORASFEKE 520
Db 1175 QKVNEKQDQNEIKKIKQIIEKNKDILTYENNIEQIEQKIELKTNAQKDDQIVNTLNE 1234
QY 521 KSKKGLKEPQOKEENSKAINN- - - - - QSG-LEEDDNITERLPEPSPIOYQQENAGLGASDPKP 578
Db 1235 VKKJIITYYKVDNQISNLVKNYEGVEYKKNVQNVNADDTDIDINDIDEINDID 1294
QY 579 YM- - - - - IKDVQVRYLAKSQIOELIKAKDYTKLAKLISNRHTYINISL- - - - - LKEOLFV 631
Db 1295 EINDIDEIKDIDHIKHFDTHFDIYHADD- - - - - TRDEYHIALSNYIKTELNI 1345
QY 632 NPR- - - - - IPSRDIKAKFVLDKTEK- - - - - NKVW- - - - - QIYSASP 665
Db 1346 NLQEIKNNIKIKFGEFSAHKEIKKESQIKNEFTKMDVYINQLDRIDRQMLDYKE- - - - - 1402
QY 666 VQNKWLSLFGYRYLLGLDPKQTIHELKVLQKAGLQPEGYNLPSDFNLEDLKNIRIKT 725
Db 1403 -LDEKYSEFN- - - - - KTKIEBNNIRENNINNVETLWYKNIIEYFLRYMNDQDKA 1451
QY 726 PLFSQ- - - - - KDNFKLSLLDFN- - - - - NYDGEIKAPEGLPLFLPKELRRSS 769
Db 1452 AKYMNIDITYKNNIBIISKQINPENYVETLKTNNMYSYVEKAND- - - - - Lfy-KQINNIII 1506
QY 770 NSGGQNSNSPW- - - - - EQEIIISQFKQNLNSQDLAQFSTKIWEIKIGDENEFDQ- - - - - NN 822
Db 1507 NS- - - - - NOLKNEAFTIDELQNIQNRKNLLTKQOIIQYTNEL- - - - - ENIFNEIKNIN 1557
QY 823 RL- - - - - QYK-LLKDLQESWINKTRDNLVYTLVGLKLVKPKNNLEAKFROISNLOELLTAF 878
Db 1558 ILVLTNYKSILODISQ- - - - - NINHVSITYEQL- - - - - HNLVYKLEB- - - - - EKEQWKTLY 1603
QY 879 YTSAAALSNMNYQDSGAKSTII- - - - - FEEIAELDPKVKKEGVADVYQKPHYAIGFDDNAGK 937
Db 1604 HKSNNVLNQNINFNEDAFINLLINIEKIKNDITHIKEK- - - - - TNYIMDVNKS- - - - - KNAQL 1658
QY 938 FNOEIVRSSRTIYLTSGSKLEADTIDQINQAVKNAPLGQSFLYDTERFQVOKLAT 997
Db 1659 YFHNTLRGKEIEYLLKNLKNSTNQOITLQELQVQENVE- - - - - KVKDIYNQTIKY- - - - - 1709
QY 998 SLAVQHQKEKTLPKKLN- - - - - NDGYTLIHDLKPKVIPQISSPEKDFWFGKLNQSQSN 1055
Db 1710 - - - - - EEKIKGNTHIITDENKINDILHNSFKQINME- - - - - SSNNKKQT 1750
QY 1056 - - - - - VNVSTFGSIIESPYFTNFQEDAD- - - - - LQDQGD- - - - - DSRGNSLND 1097
```

Db 1751 KQIIDIINDKTFEBHIKTSKTKINMLKEQSQWKHIDKTLTLEQALKLFDVINSTNNLDN 1810
Qy 1098 QEAGLLKQKLAIIQNGQFIQYQNDXIE--FEIINVEKVSLEL----- 1139
Db 1811 MLSEINS-----IQNNHTIYIQEANKSFDKFKIICDQVNDLNLKSLGLDLYNMHLKN 1864
Qy 1140 -----SFRVEFKL--AKTLEDNGKTRVLS-----DETWSLIYNTTIEKTP 1179
Db 1865 LQNEIRNNMLKFNFLDKSKIDKEBKLDILKVNISNNISLQKLYKYEALFQVK 1924
Qy 1180 MSAPVEPDTKWVEQYDPRTPLAATKTVLKFQKQIPVDGS-----GN 1222
Db 1925 KAEIQENIEKIKQINTLSDVFKPPFFI-----QLNTDSSQHEKDINNVTETKNNIDE 1980
Qy 1223 ISDKWLASIPVIHQMLRLSPVV-----KTIRELGKTEQOQQOQQOQQOQQOQ 1273
Db 1981 IYVPIQSYNLIQKYSSIFSTLNYIQTKEIKESIK-EQNLQNEKEASVLLKNIKI 2039
Qy 1274 -----KAVRKBEELETYNPKDEFNINPLTKAHLRLTSLNL-----VNNDPNYKIEDL 1320
Db 2040 NETIKLPQIKNERQNDVHNKEDYNLLQYLYNMYKMEQKLYKNDVHMDKNY----- 2094
Qy 1321 KVINEAGDHQ-----LAFSLRANNIKRLMNTPIITADYNPFYFYNEDWRSIDKYLNN 1373
Db 2095 --VENNGEKEKLLKBTISSYDKINNINNKL-----YIYKKE-----DTYFNN 2137
Qy 1374 KGNVS-----SHQOQAGNGQSGLIQRLNKNIKPFTPTPALIAIKDRNNTNL 1421
Db 2138 MIKVSLEILLNIIKKQNEQRIVINAEYDSSLINK-DEEIKKE-INNOIIEI-LKNHNEI 2194
Qy 1422 SNYSKIIMIKPKYLVERSIGVPMWSTGLDGYIGSEQTKDGTSSSSQKGFQDFQIALGL 1481
Db 2195 STIFKDIQNIKKQ-----SQQDIITNN-- 2215
Qy 1482 KNTYHGLGLSIRIFDQNE-----LAKIKDASNKGE--EKLLK----- 1520
Db 2216 --NDMHKSTILLVDIIQKKEBALNKQRLNLRNIDNINLNKKNENIIDKVIKNCDDYKDILI 2273
Qy 1521 ----SYDLFPKYNLYEYKSPKIAKGWTHIHPDQKEYPNPQKLPENY-----LNLVLN 1570
Db 2274 QNETEYQKQNLNHYEYKKSXI-----DILKIKIKQNIQKEYKNLEQMNTIIN 2324
Qy 1571 QPKVTLNYSDFITNLVPEPESGRSGTKLKQVIO-----QVNNYADWGSAY-IT 1623
Db 2325 OSIEQHFVFNADILQN-----EKILEEIIKNDILDEQIMTYHNSIDELYKLG 2373
Qy 1624 FWYDKNIITNPQNVITANIADVFIKDVKELEDNTKL 1659
Db 2374 IQCDNHLITISVVVKNKTTIMIHKKQKEDIQKI 2409

RESULT 36

Q81HY4
ID Q81HY4 PRELIMINARY; PRT; 2849 AA.
AC Q81HY4;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PP11_0392;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255705; PubMed=12369864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,

RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.,
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum".
RL Nature 419:498-511(2002).
EMBL; AB014841; AA035975.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2849 AA; 335917 MW; B551SD173D96B813 CRC64;

Query Match 3.8%; Score 373; DB 2; Length 2849;

Best Local Similarity 18.6%; Pred. No. 1.3e-06;
Matches 376; Conservative 332; Mismatches 725; Indels 592; Gaps 105;

Qy 82 KSEFTDFVSK-----PDLTNNGRVTLEIPKYQVVISSEFPDDKFRPLGLHLEKL 136
Db 403 EDEKFNDMVKNKAKDINFEMWNSKYSKISKDINL---KYKDSNDRKMSN-DSHVK--- 455
Qy 137 EDGNIAQSATKFI--YLLPLDMPKAAALGOYSIYD-----KNFNLIHP----- 179
Db 456 SSENILNNNKILNRYDLPKDI-----SYVEEYQKRKEQAKNYHNVENPNYTEDK 507
Qy 180 ----LSNFAQSISIKPLALTRSSDFIAKNQFNQDELWVY---LEKFFDLLEALKANIR 230
Db 508 NINDDKENYFGDNI-----NLNNIRSNSTTYVANSSTLSNEKXMMNLNSNIS 554
Qy 231 LQTADFSEKGNLVDPPVYFIRPNQKQEWASDLNQDKTV--RLYLRTFESFOAKTIL 288
Db 555 EDVSSIN-KENLLNTNI---NKSNETITQCCDDNNKLVNNIYIDQEF-----LI 602
Qy 289 KDY-----KYKDETFLSSIDLKASNGTSLFANEDLKD-QLDVLDDLDVDFGSGSET 340
Db 603 KYNVNVKNKDKY-DSTYIDSL-----KNNFIKDIYTDNDVIN-TESGIYNET 649
Qy 341 ITSNSQVQKVPASERSLKDRVKF-----KKDQOKPRIEKS--LYE 379
Db 650 KXNKEKISE-PCHTNNTSNIYEYLTVNDMTNVGNENSNDILNKNIEKTNFENESKKLYD 708
Qy 380 -YDALSYSQLOQLVSKPNSIKD-LVNATLARNLPSLGGKYNFL---FDDLASHDY--- 431
Db 709 VYDMINDYKNEETINKIQEKCVDKVM-----YDFINNNIDKETTNLDMDKK 758
Qy 432 YFLVSKAKIKQSSITKKLFTEL---PIKISLSSILGDOBPNIKTLFEKVTFLDNFR 487
Db 759 HLYLNNEKMNSTCDDKDTYLEMSKKEKINIFLKLQMDVN-SLSHLFQ---YFVREK 813
Qy 488 DVEIEKAFGLLYPGVNEELQARKAQRASPEKSKKGLKEFSQKSEENSKAINNQEGLE 547
Db 814 DEEMKKKLQFLILGGDE-----KKQMEFMEYKINQNTQTLNKGKLGKHSVQTNNEKKIQ 867
Qy 548 EDDNITERLPENSPIQ-QQENAGLGASPKPYMIKQVQRYLAKSQIQLKAKD-- 604
Db 868 IENIIQTDIKDITKTLYIKNDMINKKTSIDSVF-FKSLSKDSYDLYNKN-KEDIKQNDTT 925
Qy 605 YTKLAKL-----LSNRHTYN-----ISLRLEQL---FDVNPRISSSDIEKAK 645
Db 926 YTKQEDMENKVDVTVNESNTETTYNEVOKINDLKVKILEKIKGCDYNTSNNNYNDSDQTA 985
Qy 646 FVL--DKTE--KNKYWQIYSSASPVFQNKWSLFGYYRYLLGLDPKQTIHELVLKLGOKAGL 701
Db 986 ILMLODKNEYSKEKYMVYNL-----IDENRNI--LSKLNDEENM 1023
Qy 702 QFEGYEN-----LPSPDFNLEDLKNIRIK---TPLESQKDNFKLSLLDF 741
Db 1024 KSNEYKKNQKRSVMVTETVFASLKSFEFERMNLKSHNERLRRIEKLYESDRIK----- 1077
Qy 742 NNYVDGRIKAPFGLPLFLPKELRRNSNSGSGNSNSPWEQEIISQFKDQNLNQDQLA 801
Db 1078 NEYIKMB-KUKESQDRLFATE--RHIEKHLNEDLNLSKKNEDMKYDLKKKNI----- 1127
Qy 802 QFSTKIWEKIIGDENEPDQ-NRLQYKLLKDLQESWINKTRDNLNWTYVLGDKLVKPKQN 860
Db 1128 -----KIIALESQIDNNLNMISNNWKNNN--NNNNDN-----NNNNN 1164

Db 943 DPEKPYKQDFENRLINKISBLDKKFKDASLXD-----HESNNGLMBEYFNNLKANLGN 997
Qy 521 KSKKGLKEFSQOKEKSAINN-----QEGLEB----- 548
Db 998 KEKTLHYEF-----DEKAVANNIKTIDINKNISNIEIETYSIYNIEEIDAIGNI 1053
Qy 549 ---DDNITERLPEN-----SPIYOQOENAGLGPDKPMYKDV---QNRQY---YLAKSQ 595
Db 1054 KLLNDVVVKVNNVNLGIGKEKL-----KRYNPLDFGKEENIKYINEIKIKND 1103
Qy 596 IOELIKAKD-----YTKLAKLLSNRHTYNISRLK---BOLFVDPNRPSPSSRDIE-KAKFV 647
Db 1104 INTLQOXKXGSIETLTKK-TSESHIGEIRGQTDKLEKADINTYXEDPKIEKKIENV 1162
Qy 648 LDKTEKNK--YWOI---YSSASPFOKWSL-----FGYRYLLGLDPKQTHIELVKL 695
Db 1163 VKKXDKKNIYKEINKLNEISEKOKTSEBELKNINLSYGRSLGNIFLEQ-IDEEKK 1221
Qy 696 GQKAGLOFEGYENLPDSFNLEDLKNIRIKTLPFGQD-----NFKLSLLDFNN---Y 744
Db 1222 AERTIKAMEAY-----IEDLONIKKKSDBI-EKDMKIKWIDINEEMKALNISNDDDRN 1272
Qy 745 YDGEIKAFEPGLPLPKELARNSSNGSQNSHSPWE-QBIISQFDONLSNODQLAQF 803
Db 1273 YHTKSKDHKGISDIHDKSSKIIQNFESKESDINNINKELQENVSERKHSNDINHLSKV 1332
Qy 804 ST-----KIWEKIIDGENEFDONRL-----QVKLLKDLQE-----SWI 837
Db 1333 ENIYNILKNIKKIIDKVKETDBEIKKNSKINDELANSKILTKIKENSSTECOSKI 1392
Qy 838 NKTRDNL-----WYLGDKLVKPKNNLEAKFROISNLOE-LLTAFYV----- 880
Db 1393 ESTIDDDYISKIKDIADLKTYI-----LSENNINTYLNKAENYNNVLLNFHNIEMGD 1447
Qy 881 -----SNAALSNWNY-----YODSGAKSTIIFEBIAELDPKVEKV 916
Db 1448 TKSQYILNKKNSGTNSDYNINELKEHKESNGYKDEAGKNTETIKKNEKLFERYKQDV 1507
Qy 917 GADVQLKPHYAGFDDNAGKF---NOBVIR---SSRTIYLTSGSKLEADTI----- 965
Db 1508 ---TVLLNKYAVELKNFKDKTIDSEQIIKEIKEAHTNCISQSGSEQKQNEIKNEKIH 1564
Qy 966 -----DOLNQAVERNAPLGLQSF-----YLD 986
Db 1565 IEDEVANNKSKATINLKLSVDQETRFKINDIKKKSDECLKETENIENKISTLSIDT 1624
Qy 987 -----ERFVQKATSLAVQHK-----QKEKT--LPKLAN-----DGYTL- 1021
Db 1625 QETRLKENGIEFTLKLKFLSKLQKKNQKVIDDQKKELOGVNSKIENIESVNNQHKKNYEIG 1684
Qy 1022 IHDKL-----KKPVIPOI-----SSSPEKD-----WPEGKLNQ-- 1050
Db 1685 IVEKINETAKANKDOUESTKLIITPIETIENLTSSPKTDLGGDTXENLGTYNKEMNTYN 1744
Qy 1051 -----GOSQNVNVSFTGSIIES-----PYFS----- 1071
Db 1745 EFMESYNLIVDYLETVSKESITYGIEKNKRISTQKELLKSIENVNNAKSYLDDIETNEFD 1804
Qy 1072 ---TNFOEDA-DLDQDQODDSRQGNNSLDN-----QEAGLL-----KQKLA 1109
Db 1805 RIVTHFKKINDVNDKFTNEYSKVNKGFDNLSINNVKKSTDENLILLSINQTKEMVAN 1864
Qy 1110 LIGNQIOFY-QO-----QNDKEIEF-BIINVEKVSSELSFRVEFKLA 1148
Db 1865 IVSKKYYSKYEAENIFINISKLANSNIQIQNSSGIDLHKNNINIAIVSYLNSQKEDMLS 1924
Qy 1149 ---KTLEDNGKTRVLSDEMSLI-----VNTTIEKTPEMSAVPEVPTDKWQYQDPTPLA 1202
Db 1925 FIPSPQKTSYETVKKISDSYNTLLDIFFKKSXLQK-KEQALNLIFENRLLHD----- 1975
Qy 1203 AKTKFVLAFKDOIIPVDGSGNLSKWLASIPLVIHO--QMLRLS-----PWVKTIREL 1253
Db 1976 -KVQATNELKDTL---SDLKNKKEQILNKVXLLHLHKSINELNKLSCSNQNYDTILESSYDK 2032

Qy 1254 LKTEQOQOQOQOQO-----QOOPQKKAFAKEBELET---YNPKDEFENIL-----NPL 1297
Db 2033 IKESNNYEQKNGKLGIDPDVTAMEKFFNNDIKDIELENNYKHSDKDNYPSENNIL 2092
Qy 1298 TKAHRL-TLSQNVNPNYKIBDLKVIKKNAGOHQALAFSLRANNIKELMNT----- 1347
Db 2093 QSKKJLKELSNAPNAEIK-KXEDIIEKN-----GLINKLIETREKCEMFTY 2138
Qy 1348 -----PITFADYNPFF-----YINEDWRSIDKYLNN-----KGNVS 1378
Db 2139 TTLVEALRIKITYSKFITSATFSKEFLKYIDTNSLNDNDINTLOTKYDLNQIKKHTV 2198
Qy 1379 SHOQOAGNGQSGLIQRLNKNIK-----PETFPALIALKDRNNTNLS-NYSKDIIMIK 1432
Db 2199 SMPADAT--NDNNNLIEKEKEATKTINNLTELFT-----IDSNIDADVLHNNKIQMI- 2249
Qy 1433 PKYLVERSIGVWSTGLDYGISEQTKDGTSSSSQKGFQDPDFOALGLKNTYHGLGL 1492
Db 2250 -----YFNSSELHKSIDSIKQLYKQKHVFKLLNIGHINKKY----- 2284
Qy 1493 SIRIFDPGNELAKIKOASNKKGEKLLKYDLFKVYLNVEYKKSPIAKGWTNIHPDQKE 1552
Db 2285 ---FDISKEDNLLQOSES-----LKENLNDLKKIGOKISDKKNKFNELSE 2329
Qy 1553 YPNPN-OKLPENY-----LNLVLKQPKVTLNYSDFITNLFVPEPESDRGS 1598
Db 2330 IPIPNFTLKEIYHEIVKYESQIDEIKNITBENDNITLY--MDIITKLM----- 2377
Qy 1599 GTKLKQVIOQVNNVADWGSAYLTFWYDKNIT---NQPNVITANITADVPIDVKELED 1655
Db 2378 -KKVESIL-----NFTVTYENDSNVIKQHIQDNNENDVSKIKD 2414
Qy 1656 NTKLIAPNITQWPNPISGSKEKFKPTVPFGKVENENSSMSQAQTPTEWKI----- 1707
Db 2415 NLKTTIESFOKILNKLEIKAQFY-----DNNINNVISTISQDVIDVKKHISK 2463
Qy 1708 -----REGFALQ-----ALKSSFDQKTRTPVLTNAPLPLWKYGLFGFQNGPNKPTQDWR 1757
Db 2464 DLTXENELIBIQKSLEYIKKSTVDIRSEQITKYVN---PIHDY-----VEQQT 2508
Qy 1758 LVQNDN--QIAALRVOEQDRPEKSEDKQKWKFKVPIPEEMNSGNIRFVGWQI 1815
Db 2509 KKIQNDPNDKDEIDDL-IQEIYVNYNKSELSK-----LPTIINNKONV----- 2548
Qy 1816 QGPNTLWLPVINS-SVIYDFYRGTDGSDNVANLVA 1850
Db 2549 -----TPILSRIDKVINLIKSEYNNNDNVSYNVA 2577

RESULT 38

Q7RBW8 PRELIMINARY; PRT; 2070 AA.
AC Q7RBW8;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Rhopty protein (Fragment).
GN Name=PY06018;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiolini S.V., Suh B.B., Kooij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallow S.J., van Aken S.B., Riedmuller S.B., Feldblum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,


```
QY 1107 LAI-LLGN---QFIOYQONDKEIEFEIINVKVSELSFRVFEKLAKTLEDNGKTIRVLS 1162
Db 1917 SGIDLFKINIAILPYLDSQKQDTLTFTIPSEKTSYTKI-----SDSYNTLLDIL 1969
QY 1163 DETMSLIVNTTIEKTPENSAPVEPDTKWVEQDPRTPLAATKTVLKFQDQIPVDGSGN 1222
Db 1969 KRSQEL-----QKQEQALMLIFENRLHD-----KVQATNELKDTL--SDLKN 2010
QY 1223 ISDKWLASIPVHQ--QMLRLS-----IRELGKTEQOQQO-----1262
Db 2011 KKEQILNKVYKLLHKSNEKLKLSQNSQYDTILESSKYDKIKKSNYKEKENLGINFD 2070
QY 1263 -QOQOQOQOQOQKAVKKEBELTYNPKDF-----NILNPLTKAHLRLTSLNVLNNDPN 1314
Db 2071 VRAMEEQFNNDIKIEKLENNYKSKENYNFSEENNNILOSKKLELT--NAFNAETK 2128
QY 1315 YKIEDLVKIKN-----EAGDHQAFSLRANNIKRLMNT-PITPADYNPFF-----YVN 1361
Db 2129 -KIEDKILIEKNGLINLKIETKDCMLFT-----YKTLVETLIKITYDTYKFTSATKSK 2182
QY 1362 EDWRSIDKYLNNKGNVSSHQQAAGNQGSLIORLNKNIKPEPTFPALIALKDRNNTNL 1421
Db 2183 EPLKYIDATSNLNDINTLQTKYDLNQ-----INKHV-----ASMVADATNDNNN- 2228
QY 1422 SNYSDKIIMKPKYLVERSIGVPWSTGLDGYIGSEOTKDGTSSSSQOQGFQDQFQALGL 1481
Db 2229 -----LIEKE-----KEATKT--INNLTFTIDSNNKIDADGL 2259
QY 1482 KNTYHGKLGSLTRIFDPCNELAK---TKDASNNKGEEKL-----KSYDLFQNY-- 1528
Db 2260 HNNK-----IQIIFYNSELHKSIDSIKQYKQKQFHLALSETPIPEFNWTLKEIYHDIYKYNQID 2311
QY 1529 -----LNEYEKSPKIAKGTWNIHPDQKEYPNP--QKLPENY-----1564
Db 2312 ILQLOBSELTAANLNDLKEIGQISDKKQFLHALSETPIPEFNWTLKEIYHDIYKYNQID 2371
QY 1565 --LNLVLNQPKVTLNNSDFITNLVFEPEGSDRGSGTKLKQVIOQVNNVADWGSAYL 1622
Db 2372 EIEENITNEENITLY--MDII-----TKLMKKVQSILN-----EV 2405
QY 1623 TFWYDKNIITNQPNVITANITADVFIDKVELEDNTKLIAPNITOMWPNISGSKPKYKPT 1682
Db 2406 TTY-----ENDSNIILKQIHDNNENDVSKIESLETTTQSFQKILNKLINEKAQFY--- 2456
QY 1683 VFFGNWENSSMNSQAQPTW-----EKIREGFALQALSKSPDQKTRTFVLTT 1731
Db 2457 -----DNNNNINNVISTISQDVNDVKKHSKDLTIENELIQIKSLEDIKKSTYDIRS 2508
QY 1732 NAPLPLWKY-GPL-----GFQGNPKPTQDWRLVFNQDNDNQIAALRVQODRPEKS 1781
Db 2509 E---QITKYVNPPIHDYVEQOTKKIQNNPN---KD-----EIDD-----LIQBIVNNYKE 2551
QY 1782 SEDKQKWKIKFKVVPPEEMFNSGNIRFCVGMQIQGPNTLWLPLVINS-SVIYDFYRGTD 1840
Db 2552 SELK-----LPTIINNKNV-----TPIISRIDKVINLIKSEYN 2585
QY 1841 SNDVANLNAPWQVKTIAFTNNAFNPVKEFNISKKIVE 1879
Db 2586 NNDNVSTNVAK---KLBEDANNIIRDITSHNMLNDLIQ 2621

RESULT 40
Q25662 ID Q25662 PRELIMINARY; PRT; 1939 AA.
AC Q25662
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Repeat organellar protein.
OS Plasmodium chabaudi.
OC Eukaryota, Alveolata, Apicomplexa, Haemosporida, Plasmodium.
OX NCBI_TaxID=5825;
```

```

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=96V;
RA Werner E.B., Holder A.A., Aszodi A., Taylor W.R.;
RT "A novel 11-residue coiled-coil motif predicts a histidine zipper.";
RL protein rept. Lett. 3:139-146(1996).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=96V;
RX MEDLINE=98418765; PubMed=9747969; DOI=10.1016/S0166-6851(98)00067-X;
RA Werner E.B., Taylor W.R., Holder A.A.;
RT "A plasmodium chabaudi protein contains a repetitive region with a
predicted spectrin-like structure.";
RL Mol. Biochem. Parasitol. 94:185-196(1998).
DR EMBL; U43145; AAC63403.1; -.
DR PIR; T18372; T18372.
DR HSSP; P13848; INO4.
SQ SEQUENCE 1939 AA; B36E462001C6F22F CRC64;
Query Match 3.8%; Score 370.5; DB 2; Length 1939;
Best Local Similarity 19.3%; Pred. No. 1e-06;
Matches 394; Conservative 325; Mismatches 683; Indels 643; Gaps 98;
QY 68 SVKALLN--GKTFDPKSSFTDFVSK-----PDFLTNNGRTVLEIPKXYQVVISFSS 118
Db 4 NLKSKKNEDSGNSKDKTNETSGIEKKEKSNKYKYNVNSTKKDKDNDSIVYDDES 63
QY 119 PEDDKERFLGFLHKE-KLEDGNTAQSATKTYLLPLDMPKAAALGOYSIVYDKNFNNLI 177
Db 64 KVGENDH-----HMKEYELEQD--LKETLK3ITALSIKV-----EYEVKIEELEKELKL 111
QY 178 HPLSNFSAQSIKPLALTRSSDPI-----AKLNQFNQDDELWV-----214
Db 112 EKEQINKEYEK--ELNEKSEFIKQOMELLKEKELNINLKENKINNKEIITLKREKLNLD 169
QY 215 ---YLEKFFOLEALK---ANIRLQ---TADFSPEKGNL---VDPFVYSFTRPNQOKEW 261
Db 170 TSEYIEKNKEKEKLYNEVTNIKMSLDKLTCEVQEKDNLEKINKKVKIEKENNRELKEF 229
QY 262 ASDLNQ-----DQKTVRLYLRTEFSQAQKTLKDYKYKDEFTFLSSIDLKASNGTS 311
Db 230 MKERKEIIESLGTINDKNAYEKLEISFEKRRMI-----265
QY 312 LPANENDLKQDLDVLLDVSDYFGQSETITSNSQVKKVPVASE---RSLKDRVKFKKQDQ 368
Db 266 ---EMLDSKLIKEENF-----ANKQAKLEKENEIIEKLD--IESREKDF 307
QY 369 KPRIEKSFLYEYDALSPYSQ-----QELVSPKNSIKOLVATLARN 410
Db 308 KSKEEKFASMENELNTLSKDLKSNACQMEVYKLEIKDLSQSLVEKEREIFEIKNE-----362
QY 411 LRPSLGKYNFLPDDDLASHLDYFLVSKAKIKOSSITKKLFIELPIKISLKSSILGDOEPN 470
Db 363 ---YDD-----KINNMKELSSINDK-----GIDNTVLHSEEEK 393
QY 471 I-KTLFEKVTFKLONFRDVEIEKAFGLLYPGVNELEQAKAQRASFEKSKKKG-----525
Db 394 INKLLKEKETELN-----EIHKKYNLEIETIKNELNEKEE-----ELEKNKRAHTVE 440
QY 526 ---LKEPS--QOKBEENK-----AINNQEGL--EEDDNITERLPENSPIQOQENA 569
Db 441 VTNLTKEIKLEKKTDEKAGEGHKNELNELNQLSKLNKEKDN-----KNENT 488
QY 570 GLGASDPKPYMIKDVQONQRYYLAKSQIQELIKAK-----DYTKLAKLNSN-RHTYNIS-- 621
Db 489 EL-----NDKISSINSEVNTLNKDKQTLGNDIKTLNDLNLNKNINNTSDN 534
QY 622 --LRLKEQLFDVNPRIIPSSRDIEKAKFVLDKTEKKNKYQIYSSASPVFQNKWSLFGYRY 679
Db 535 KNNKMKEDLAMLN-----EEMEGKCVWIDELEK-KY-----564
QY 680 LLGLDPKQTIHEL-VKLGQKAGLQFEGYENLPSPDFNLEDLKN-IRIKTPLFSQKNPF--- 734
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2005, 20:54:27 ; Search time 60 Seconds
(without alignments)
3013.188 Million cell updates/sec

Title: US-10-607-631-8
Perfect score: 9732
Sequence: 1 MNKKSTLLATAAIIIGST.....TNNAFNVKFNISKIVE 1879

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	394	4.0	1099	2 T18357	mhl protein - Myc
2	393	4.0	1092	2 T18354	adhesin - Mycoplas
3	391.5	4.0	1108	2 T18353	protein p97 - Myco
4	377	3.9	2269	2 T28677	thoptry protein -
5	370.5	3.8	1939	2 T18372	repeat organellar
6	366	3.8	2401	2 T28676	thoptry protein -
7	340.5	3.5	3394	2 T18501	hypothetical prote
8	333	3.4	2166	2 G70163	hypothetical prote
9	331.5	3.4	1738	2 T14867	interaptin - slime
10	325	3.3	4688	2 F82885	hypothetical prote
11	319.5	3.3	2819	2 A90551	conserved hypotet
12	312	3.2	2244	2 F90563	hypothetical prote
13	305.5	3.1	1875	2 S38173	myosin-like protei
14	301	3.1	1631	1 SAZQK1	major merozoite su
15	301	3.1	2663	1 S28261	centromere protei
16	295	3.0	1639	2 S05603	major merozoite su
17	291.5	3.0	1727	2 T50073	myosin-like coiled
18	291	3.0	1640	2 A24594	probable major sur
19	291	3.0	5005	2 F82884	hypothetical prote
20	290.5	3.0	1679	2 S48385	hypothetical prote
21	290.5	3.0	1979	2 C71622	hypothetical prote
22	289.5	3.0	1302	1 JC6009	surface-located me
23	289.5	3.0	3848	2 T17414	tipC protein - sli
24	287	2.9	1183	2 F90559	conserved hypotet
25	282.5	2.9	2829	2 A42771	reticulocyte-bind
26	281	2.9	1805	1 A64224	hypothetical prote
27	281	2.9	2441	2 D71623	erythrocyte membra
28	280.5	2.9	1818	1 S73852	hypothetical prote
29	279.5	2.9	2231	2 D71870	hypothetical prote

30	279.5	2.9	3216	2 C90538	hypothetical prote
31	278	2.9	3724	2 T18427	hypothetical prote
32	276.5	2.8	1690	2 T13030	microtubule bindin
33	276	2.8	3259	1 A56539	giantin - human
34	274.5	2.8	1997	2 F71607	DNA helicase II BR
35	272	2.8	1516	2 E71619	RAD2 endonuclease
36	271	2.8	3225	2 I52300	giantin - human
37	270.5	2.8	1365	2 T30822	impl protein - Myc
38	270.5	2.8	2599	2 F90608	ABC transporter pe
39	270	2.8	1127	2 T28317	ORF MSV156 hypotet
40	267	2.7	1830	2 E82909	conserved hypotet
41	264.5	2.7	1088	2 T18559	hypothetical prote
42	264.5	2.7	1208	2 T39068	coiled coil protei
43	262.5	2.7	1701	2 A54498	major merozoite su
44	262.5	2.7	2295	2 B71621	probable membrane
45	261.5	2.7	2178	2 S55805	alpha-toxin - Clos

ALIGNMENTS

RESULT 1

T18357

mhl protein - Mycoplasma hyopneumoniae

C:Species: Mycoplasma hyopneumoniae

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T18357

R:King, K.W.; Faulds, D.H.; Rosey, E.L.; Yandey Jr., R.J.

Vaccine 15, 25-35, 1996

A:Title: Characterization of the gene encoding Mhl1 from Mycoplasma hyopneumoniae and ex

A:Reference number: Z18890

A:Accession: T18357

A>Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1099 <KIN>

A:Cross-references: UNIPROT:Q49560; EMBL:U27294; NID:g1403588; PID:g1403589; PIDN:AAB034;

C:Genetic:

A:Genetic code: SGC3

A>Note: mhl1

Query Match 4.0%; Score 394; DB 2; Length 1099;
Best Local Similarity 20.5%; Pred. No. 3.8e-09;
Matches 261; Conservative 213; Mismatches 475; Indels 326; Gaps 59;

QY	1	MNKKSTLLATAAIIIGSTVFGTGLASVKYRGVNPQTQGVISQLGLDSVAPKPSIA	60
DB	1	MSKSKTFKIGTAGIVGVLGTVGLSSLAKYRSESPRKANDPAKVSTLAFSPYAF	60
QY	61	NFTSDYQSVKALLNGKTFDPKSSSEFTDFVSKFDLTNNGRVLEI-----PKYQV	112
DB	61	ETDSDYKIVKRWLVDSNN-NIRNKE--KVIDSFPTKNGDQLEKINFQDPYTKAKITF	117
QY	113	VISEFSPDDKFRIGLGHKLKEDGNTAQATKFIYLLPLDMPKAA---LCQYSYVD	169
DB	118	EILEIIPDDVNQNFVKVFOALQKLNHGDIKSD---IYEQTVAFAKOSNLLVAEFNFSLK	174
QY	170	KNFNNL-----IIHPLSNFS---AQSIKPLALTRSSDFIAKQNFNNODELVYLEKFF	220
DB	175	KITEKLNQIENLSTKITNFADEKTSQKDFSTLRAIDFQYDLNTARNPEDLDIKLANYF	234
QY	221	DLEALKANT-RLQTADFSPKGNLVDPPFVYGFIRNPQNKQEWASDLNQDKTVRLYLRT	279
DB	235	PV--LKNLNLNNAENKLPNNLGNIFKFGPAKD--SSTNQVYSIQNQ---IPSLFLKAD	288
QY	280	FSPQAKTILKDYKDET--FLSSIDLKASNTGTSLSFANENDLKQDLVDLLVSDIFGGQ	337
DB	289	LSQSAREILLAS---PDEVQPVINILRLMKDNSSSYFLNFEDFYNNL-----	331
QY	338	SETITSNQKVPVPAESRSILKDRVKKQDQKQRIEKFSLYEY-----DALSFY	386
DB	332	-----TLKN---MOKEDLNKQNLNLSAYFLADIKSGFFGDKRSSH	370
QY	387	--SQLQELVSKPNSIKDLVNLARNLRFSLGKYNFLFDDL--ASHLDYYFLVSKAKIKQ	442

Db 371 TKSEISNLNKKENIYD-----FKYNGKFNDRLNSPNLEYSILDAASAQLDK 417
Qy 443 SSITKKLFIELPKISLSSILGDO-EPNIK-TLFEKEVTEFKLONFRDVEIEKAFGLLYP 500
Db 418 KD---KSIIILPYRLEIKKFFADDDLYPTDKNVLVKEGILKLTGFK-----KGPKIDLP 469
Qy 501 GVNELEOAKRAQASFEKSKKGLKEFSQOKEENSKAINNQEGLEBDDNITERLPENS 560
Db 470 NINQOI-----FKTEY---LPFFEKGEEQAKL-----DYGNI----- 500
Qy 561 PIQYQENAGLSPDKPYMIKDQVQRYYLAKSQIOELIKAKDYTKLAKLLSNRHTYNI 620
Db 501 -----NPNYNTQ-----LAKVEVEALFKGNKQEIYQALDGNAYEYF 536
Qy 621 SLRLKEQLFDVNPRI--PSSRDIEK-----AKFVLDKTEKNKYQIYSS--ASP 665
Db 537 G-AFKSVLNSWTGKIOHPEKADIQRTHLQSVKIGNSVLNQPTTKB-QVISLSKSN 594
Qy 666 VFQNKWSLFGYRYVLLGLDPKQTHIELVKLGQKAGLQEGYENLPSDFNLBDLXNI----- 721
Db 595 FFKNGHVASYPQDILLTKDKLTVLETLVYLAKKWLGE-TNWAQPKG-AFYTKDIPAE 652
Qy 722 -RIKTPLESQDNF-----KLSLLDPNNYYDGE-IPAEFGLPLFLPKELRRNSNSGGSQ 775
Db 653 DKLFLEWKKKDPYVQINIEIHLQSFNILARNDVIKSDGYGVLLPQSVKTELEGKN--- 709
Qy 776 NSNSPWEQELISQPKDQNLNQDQLAQFSTKIWEKIGDENEFDQ--NNRLQYKLLKDLQ- 833
Db 710 -----EAQIFEALKYSLIEN---SAFKTTILOKNLLEGDTFTFGDFLAKFLLKAAQF 760
Qy 834 ---ESWINKTRDNLWYTY---LGDKLKVKPKNNLEAKFRQISN-LOELL----- 875
Db 761 NNFAPWA-KLDDNLQYSFEAIKKGTTKEGKREEVKKVKELDKNKIGILPQPPAAKPE 819
Qy 876 ---TAFYTSAAALSNWNYQSGAKSTIIFPEIAB-----LDPKVEKVGKA- 918
Db 820 AKPVAAPVAA-----KPEAAKPVAAKPEAAKPVAAKPEAAKPVAAK 871
Qy 919 -----DVYQLKFAVIAIGFDDNAGFNQOEIVIRSSRTIYLTSG 956
Db 872 PVATNTNTWGTSLNKNPKEDIPPMAPFSKLEYTDE-----NKLSTKTPINVFELVH 927
Qy 957 KSKLE-ADTIDOLNQAQVKNAPLQSFYLDTERFGVQFKLATSIAVQHKQKERTLPKKLN 1015
Db 928 QSEVEDQKIIKELDKTVLNLQYQFQEVKVTSDQ---YQKLSHPWMTGCTQNGKKGEPT 984
Qy 1016 NDGYTLIHLKLLKPKVPIQISSPEKDWPEGLNQ---NGQSNVNVSTF-----GSI 1064
Db 985 NQG-----KKAGAPSGQKKAGSTPNQKKASGTPNQOSPTSELTYNLPDLGKKIDEI 1037
Qy 1065 IESPVFTSNFQEDADLDQD-----QDSSROGNSLNDQAGLLKQKLAILLG 1112
Db 1038 IKKO--GKNWTEVELIEDNITAGDAKLLYFLTRDDSKG-----DPKSSLKVKIITVQS 1090
Qy 1113 NQFIQYQONDKEIE 1127
Db 1091 N-----NNQELK 1097

RESULT 2

T18354
adhesin - Mycoplasma hyopneumoniae
C:Species: Mycoplasma hyopneumoniae
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18354
R:Wilton, J.L.; Scarnan, A.L.; Walker, M.J.; Djordjevic, S.P.
Microbiology 144, 1931-1943, 1998
A:Title: Reiterated repeat region variability in the ciliary adhesin gene of Mycoplasma
A:Reference number: Z18887; MUID:98361039; PMID:9695926
A:Accession: T18354
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A;Residues: 1-1092 <WIL>
A;Cross-references: UNIPROT:O07132; EMBL:AF001398; NID:g2150123; PID:g2150124; PIDN:AAC33
C;Genetics:
A;Genetic code: SGC3

Query Match 4.0%; Score 393; DB 2; Length 1092;
Best Local Similarity 20.6%; Pred. No. 4.2e-09;
Matches 261; Conservative 204; Mismatches 464; Indels 340; Gaps 59;

Qy 1 MNKKSTLLATATAAIIGSTVFGVGLASVKVYGVNPTQGVISQLGLDSVAPKPSIA 60
Db 1 MSKSKTKFKLGLTAGIVGLGVGLTVGLSLAKTRSRSPRKIANDFAAKVTSLAFSPYAF 60
Qy 61 NFTSDYQSVKALINGKTFDPKSEFTDFVSKFDFLTNNGRTVLEI-----PKYQV 112
Db 61 ETDSDYKIVKRWLVDSSNN-NIRNKE--KVIDSFSEFTKNGDQLEKINFQDPPEYTKAKITP 117
Qy 113 VISEFSPEDDKERPRLGPHLKEKLEGNIAQSATKFIYLLPLDMPKAA---LGQYSIVD 169
Db 118 EILEIIPDDVQNQFVKVFAQLQKLHNGDIAKSD---IYEQTAPAKQSNLLVABFNFSLK 174
Qy 170 KNFNNL-----IHLPLSNFS---AQSIKPLALTRSSDFTAKLNQFNQDELWYLEKFF 220
Db 175 KITEKNQOIENLSTKITNFADKTSQKDPSTLRAIDFQYDLNLTARNAEDELKLANYP 234
Qy 221 DLBAALKANI-RLQTADFSFEKGNLVDPPVYSFIRNPQNKQEWASDLNQDQKTVRLYLRT 279
Db 235 PV--LKNLNLNAPENKLPNNLGNIFEFSPAKD--SSTNQYVSIQNG---IPSLFLKAD 288
Qy 280 FSOAKTILKDYKYKDBT--FLSIDLKAAGTSLPANENDLKDQDVLDDVSDYFGGQ 337
Db 289 LQSAREILAS---PDEVQVVINILRMKKDSSYFLNFDVNNL----- 331
Qy 338 SETITNSQVQPVASERSLKDVKPKDOQKPKRIEKFSLVEY-----DALSFY 386
Db 332 -----TLKN---MOKEDLNKAGQNLISAYEFLADIKSGFFPDGKSSH 370
Qy 387 --SOLOELVSKPNSIKDLVNNATLARNLRFSLGKYNFLFDDL---ASHLDYFYFLSKAKIKQ 442
Db 371 TKAISNLLNKKENIYD-----FGKYNKGFNDRLNSPNLEYSILDAASASLDK 417
Qy 443 SSITKKLFIELPKISLSSILGDO-EPNIK-TLFEKEVTEFKLONFRDVEIEKAFGLLYP 500
Db 418 KD---KSIIILPYRLEIKKFFADDDLYPTDKNVLVKEGILKLTGFK-----KGPKIDLP 469
Qy 501 GVNELEOAKRAQASFEKSKKGLKEFSQOKEENSKAINNQEGLEBDDNITERLPENS 560
Db 470 NINQOI-----FKTEY---LPFFEKGEEQAKL-----DYGNI----- 500
Qy 561 PIQYQENAGLSPDKPYMIKDQVQRYYLAKSQIOELIKAKDYTKLAKLLSNRHTYNI 620
Db 501 -----NPNYNTQ-----LAKVEVEALFKGNKQEIYQALDGNAYEYF 536
Qy 621 SLRLKEQLFDVNPRI--PSSRDIEK-----AKFVLDKTEKNKYQIYSS--ASP 665
Db 537 G-AFKSVLNSWTGKIOHPEKADIQRTHLQSVKIGNSVLNQPTTKB-QVISLSKSN 594
Qy 666 VFQNKWSLFGYRYVLLGLDPKQTHIELVKLGQKAGLQEGYENLPSDFNLBDLXNI 725
Db 595 FFKNGHVASYPQDILLTKDKLTVLETLVYLAKKWLGE-----ETNRAQFPKVEYFTK--- 646
Qy 726 PLFSQDNFK-----LSLLDPNNYYDGE-IPAEFGLPLFLPKELRRNS 769
Db 647 DIFAEADKLFLEGGKKDPYVQINIEIHLQSFNILARNDVIKSDGYGVLLPQSVKTELE 706
Qy 770 NSGGSQNSPWEQELISQPKDQNLNQDQLAQFSTKIWEKIGDENEFDQ--NNRLQYK 828
Db 707 GKN-----EAQIFEALKYSLIEN---SAFKTTILOKNLLEGDTFTFGDFLAKF 754
Qy 829 LKDLQ-----ESWINKTRDNLWYTY---LGDKLKVKPKNNLEAKFRQISN-LOELLTAFY 879
Db 755 LKAAQFNNFAPWA-KLDDNLQYSFEAIKKGTTKEGKREEVKKVKELDKNKIGILP--Q 811

```
QY 880 TSAALSNWNYQDSGAKSTIIFEEIAELD-----PKVKEKYCA----- 918
Db 812 PPAA-----KPEAAKPVAAKPEAAKPEAAKPVAAKPVAAKPVAAKPVATNTN 863
QY 919 -----DVQLKFHYAIGFDDNAGFNQEVIRSSRTIYLKTSKSKLEAD-T 964
Db 864 TWYGFSLTNKPKEDYFPMAPSYKLEYTDE-----NKLSLATPEINVLFLVHQSEVEEQKI 919
QY 965 IDOLNOAVKNAFLGLQSYFLDTERFGVFOKLATSLAVQHKQKEKTLPPKLLNDGYTLIHD 1024
Db 920 IKELDKTVLNLQYQFQEVKVTSEQ-----YQKLSHPMMTEGSPNQGKKAEGAPNQ----- 970
QY 1025 KKKKPV-----IPQISSPEKDMFEGKLNQ-----NGOSQNVNVSTF-----G 1062
Db 971 --KKAEGAPSQGKKAEGAPNQGKKAEGSPQGKKAEGASNOQSTTTELNTLYLPELKKID 1028
QY 1063 SIIESPVSTNFOEDADLDQD-----QDSSRQGN-----NSLDN 1097
Db 1029 EIIKKQ--GKNWKTVELIEDNIAGDAKLLYFVLRRDSSKSGDPKSSVKVKITVQKSNNN 1086
QY 1098 QEAGLLKQK 1106
Db 1087 QE---LKSQ 1092

RESULT 3
T18353
protein p97 - Mycoplasma hyopneumoniae
C:Species: Mycoplasma hyopneumoniae
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18353; T03805
R:Hsu, T.; Artushin, S.; Minion, F.C.
J. Bacteriol. 179, 1317-1323, 1997
A:Title: Cloning and functional analysis of the p97 swine cilium adhesin gene of Mycoplasma hyopneumoniae
A:Reference number: Z18886; MUID:97175562; PMID:9023217
A:Accession: T18353
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1108 <HSU>
A:Cross-references: UNIPROT:Q49542; EMBL:U50901; NID:g1399525; PID:AAB478
R:Hau, T.; Minion, F.C.
Gene 214, 13-23, 1998
A:Title: Molecular analysis of the p97 cilium adhesin operon of Mycoplasma hyopneumoniae
A:Reference number: Z15097; MUID:98391007; PMID:9729120
A:Accession: T03805
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1089-1108 <HS2>
A:Cross-references: EMBL:AF012905; NID:g2654768; PID:AAC32526.1; PID:g2654769
A:Experimental source: strain 232A
C:Genetics:
A:Genetic code: SGC3
A>Note: P97

Query Match 4.08; Score 391.5; DB 2; Length 1108;
Best Local Similarity 20.2%; Pred. No. 5e-09;
Matches 258; Conservative 210; Mismatches 486; Indels 325; Gaps 57;

QY 1 MNKKSTLLATAAIIGSTVFGTVGLASKVKYRGVNPQTGVISQLGLDSDVAFKPSIA 60
Db 1 MSKSKTFKGLTAGIVGLGVGLTVGLSSLAKYRSSEPRKIANDFAAKVSTLAFSPYAF 60
QY 61 NFTSDYOSVKALLNGKTFPPKSSSEFTDFVSKDFLTNNGRVTLEI-----PKYQV 112
Db 61 ETDSDYKIVKRWLVDSNN-NIRNKE--KVIDSFSFTTKNGDQLEKINFQDPPEYTKAKITF 117
QY 113 VISFSPEDDKERFLGHLKLEKLEDCNIAQSATKFTYLLPLDMPKAA---LGQYSYIVD 169
Db 118 EILEIIPDDVNQNFVKVQALQKLNHGDIAKSD---IYEQTAFAPAKOSNLLVAEFNFSLK 174
QY 170 KNFNNL-----ITHPLSNFS---AQSIKPLATRRSDDFTAKLNQFNQDELWYLEKFF 220
Db 175 KITEKLNQOENLSTKINTFADEKTSQKQPSLRAIDFQYDLNTARNPEDLDIKLANYP 234
```

```
QY 221 DLEALKANI-RLQADTFSEKGNLVDPPVYGFIRPNQONKEWASDLNQDQKTVRLYLRT 279
Db 235 PV--LKNLINRNAPENKLPNNLGNIFEFAKD-SSTNQVVSIONQ---IPSLFLKAD 288
QY 280 FSPQAKTILDKYKQDET--FLSSIDLKASNGTSLFANENDLKDQDLVDLLDVSYFGQ 337
Db 289 LSQSAREILAS--PDEVQPVNILRLMKONSSYFLNPFEDFVNVL----- 331
QY 338 SETTITSQVFPVPAERSLSKDRVKFKDQKQPKRIEKFSLYEY-----DALSPY 386
Db 332 -----TLKN---MOKEDLNAGQNLISAYEFLLADIKSGPPFPGRSSH 370
QY 387 --SQQLVSKPNSIKDLVNATLARNLRFSLGKYNFLPDDL--ASHLDYFVLVSQAKIKQ 442
Db 371 TWAEISNLNKENIYD-----PKYNGKFNDRLNSFNLEISLDAASASLDK 417
QY 443 SSITKKLFIPIKISLSKSSILGQ-EPNIK-TLFEKEVTFKLDNPRDVEIEKAFGLLYP 500
Db 418 KD--KSIVLIPYRLEIKKFPADDLYPDTKNILVKEGILKLTGFK-----KGSKIDLP 469
QY 501 GVNBELEQARKAQRASFEKEKSKGLKBPQSQKSEKSNKAINNOEGLEEDDNITERLPENS 560
Db 470 NINQOI-----FKTEY---LPPFEKSGEQAKL-----DYGNIL----- 500
QY 561 PIQYQENAGLGASPDKPYMIKDQVQNYVYLAKSOIQELIKAKDYTKLAKLLSNRHTYNI 620
Db 501 -----NPYNTQ-----LAKVEEALFKGNKQEIYQALDGNVAYEF 536
QY 621 SURLKEQLFDVNPRI--PSSRDIEK-----AKPVLDTKKNKQWQYISS--ASP 665
Db 537 G-AFQSVLNSWTGKIQHPEKADIQRTRHLBQVKIGSNVSNQPOPTTKE-QVSSLSKSN 594
QY 666 VFQNKWSLFGYRYLLGLDLPKQTIHELKLDQKAGL-----QF-BGYENLPDPF--NLED 717
Db 595 FFKNGHQVASYFQDLTKDKLTILETLVYLAKKWLGTNRAQFPKGVFOYTKDIFAEADK 654
QY 718 LKNIRIKTLPFSQDNF---KLSLLDFNNYDGB--IKAPEGLPLFLPKELRRSSNSG 772
Db 655 LKPLELK-----KQDPYQIKIHIHLSFNILARNVDIKSDGFVGVLLLPQSVKTELEGK 709
QY 773 GSQNSNSPWEQBIISQFKDQNLNSDQLAQFSTKIWEKIIGDENEFDQ--NNRLQYKLLKD 831
Db 710 -----EAQIFEALKKYSLIEN---SAPKTTILDKNLLEGTDFKTFGDFLKAFFLKA 757
QY 832 LQ-----ESWINTKTRNLWTY-----LGDKLKVKPKNLEAKFRQISN----- 870
Db 758 AQFNNPAPWA-KLDDNLOYSFEAIKKGGETTREGKREEDVKVKELDNKIKGILPQPPAAK 816
QY 871 ---LQELLTAFTYTSAAALSNNWYQDSGAKSTIIPETIAE-----LDPKVK 914
Db 817 PEAAKPVAAKPETTKPVAAKPEAAKPEAAKPVAAKPEAAKPVAAKPEAAKPEAAK 876
QY 915 KVGA-----DVYQKUFHYAIGFDDNAGKFNQEVIRSSRTIYL 952
Db 877 PVAAPKPEAAKPVATNTGFSLTNKPREDYFPMAPSYKLEYTDE---NKLSLKTPPEINVL 932
QY 953 KTSKSKL-EADTIDOLNOAVKNAFLGLQSYFLDTERFGVFOKLATSLAVQHKQKEKTL 1011
Db 933 ELVHQSEYEBEQEIIKELDKTVLNLQYQFQEVKVTSDQ---YQKLSHPMMTEGSSNOGKS 989
QY 1012 KKLNDGYTLIHDKLLKPKVPIQISSPEKDMFEGKLNQSQSNVNS-----T 1060
Db 990 EGTPNQ-----KKAEGAPNQGKKAEGTPNQGKKAEGAPSQOSTTTELNTLYLPLDGGK 1042
QY 1061 FGSIIIESPVSTNFOEDADLDQD-----QDSSRQGNNSLDNQEAGLLKQKLA 1108
Db 1043 IDEIIKQ--GKNWKTVELIEDNIAGDAKLLYFVLRRDSSKSG-----DPKSSSLVKIT 1095
QY 1109 ILLGNQFIQYQNDKGE 1127
Db 1096 VKQSN-----NNOEPE 1106
```

```
RESULT 4
T28677
thoptry protein - Plasmodium yoelii
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28677; C45521
R:Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
A:Title: A gene coding for a high molecular mass thoptry protein of Plasmodium yoelii.
A:Reference number: Z40508; MUID:95021522; PMID:7935623
A:Accession: T28677
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2269 <KEE>
A:Cross-references: UNIPROT:Q26223; EMBL:L27838; NID:G457145; PID:G457146; PIDN:AAA21304
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii thoptry protein. Multiple co
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: C45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2131-2269 <KE2>
A:Cross-references: GB:M34283

Query Match          3.9%; Score 377; DB 2; Length 2269;
Best Local Similarity 18.5%; Pred No. 5.6e-08;
Matches 418; Conservative 363; Mismatches 762; Indels 712; Gaps 106;

Qy 19 STVFGTVVGLASKVKYRG-----VNPTQGVISQLGLDSVAFKPSIANFTSDYOSVKKAL 73
Db 165 NTIVNTIKSYFDQI-YEGDIDTFYNELSSIVKE-DPIIDIEDTKLENLRKIDNVYDKI 222
Qy 74 LNKGTDFPKSSEFTDFVSKFPLTNGR---TVLEIPKKYQVIVSEFSP-----120
Db 223 -----QKMEIETVKSHLNNIETNNKLPNTILEI-KKY-IYDEISKELNKMLEDPKN 271
Qy 121 -----DDKERFLGLHLEKLEKEDGNIAGSATKFIYLLPLDMPKAALGOVSYIVD 169
Db 272 KEKELSNKISDYDKKREQSEVSKSMLEIRHNSQTN-----VDNTEBEAKQNY--D 323
Qy 170 KNFNLLIHLNFSQAQIKPLALTRSSDFIAKLN---QFNNDLWYLYE--RFDDI-E 223
Db 324 KSNEHMTIPTNEDBSIKIIEVTKMDEILSKVNTYIDFNKKYKETYVNSEHSQFTELTD 383
Qy 224 ALKANI---RLQADPFSEKGNLVDPFVYSIRPNQ--QKEWASDLNQDQKTVRLYL- 277
Db 384 KIKAEVSKELKCKCQSFNDK-----SLINETKNSIEKEY-QNIN-TLKKYDVEIKV 434
Qy 278 -----TEFSPOAKTILKD-----YKVKDETFLSSIDLKASNGTS 311
Db 435 CKSTKESITKFSK-QTILKMDLNLQNIKTVKTNSIDSKSYIEKEQILTGQTKLENKFT 493
Qy 312 LF-----ANENDLKQDLVDLLVDSYFGQSEITTSNQKVPVAPASERSLKDORVFKK 365
Db 494 EFLSNHNEANNELIKYFS-----DLKANLGINEENMLNYQTE-----KEKTFND- 544
Qy 366 DQKPRIEKFLSYEDALSPYSQQLVSKNSIKDLVNATLARNLRFSLGKYNFLPDDL 425
Db 545 THINEISKIBIKTH--ASIYNISEETEREIGINIESLNTKVFKEKVENVTNLNRIKEKL 602
Qy 426 ASHLDYFLVSKAKIKQSITKGL---PIELPIKLSKSSILGQEPNIKTLFE--KEVT 480
Db 603 -KHVDFSPDFGEGNIKYTDKIKKINDDIMAVSQIDQHINGLDDIQKSESIVSEMKEQI 661
Qy 481 FKLDNFRDVEI--EKAFGL-----LYPGVNEELEQAKAQAQASFEKEKS 522
Db 662 NKLEKVSNTESNDNVEGIKKKQIIVTKIDKKNIYBEINKLSEISKIEKEDNTSLEKV 721
Qy 523 KK-----GLKEFSQKSEKSKAINNOGLEED-DNITERLPE-----NSPI 562
Db 722 KDINLSYQNLGNLFLQIDBEKKAENTIKSMAYIDDLNLIKKSQIEIETEMDIKMDI 781
```

```
Qy 563 QYQOENAGLGASPDKPYMIKDVONOR-----YVLAKSQIOBELIKAKDY-----TKLAKLLS 613
Db 782 NKMEALKISHDDDKCHDKSKKHENISDIYDKSKKIIQDFRESINDIKNKLQKVVS 841
Qy 614 NRHTYNTSLRLKEQLFDVNPRISSRDIEKAFVLDKTKNKYQWQIYSSASPVPQKWSL 673
Db 842 ESQNHNSDI--NQCLNEVANIYNILKNIKKIIDVKKE-----YTSEIEKN-----886
Qy 674 FGYYRYLLGLDPKOTIHLELVKLGKAGLQFEGYENLPS-----DFNLED-----LKNIR 722
Db 887 -----KGNINDELNNSSEKVIKIEGDLSLSEKRSKINSTLDDKDIDECIKNIN 934
Qy 723 I-KTLPFSQKDNF-----KLSLLDFAN-----YYDGEIKA-----PBFGL 756
Db 935 VLKGNILNEETNITNHFKNABEYKNIVLSNFIEMADNKSQYILEIKKNGNTDHDYNI 994
Qy 757 PLFLPKELRNSNSGG-----SONSNS-PWEQBIIISQFDQNLNSQDQLAQSFSTKIWEK 810
Db 995 -----KELKSHKDKSNGYKTEADQNKKAIOKNKELFPQYKEE-----VTVLNKK 1038
Qy 811 IIGDE--NEFDQNNRLQYKLLKDLQESW-----INKT 840
Db 1039 YYAVELKNNKFDKTKNDKSKQIIKEIKDAHNYCTLESKSEKKNKNEIKHIEDEVANN 1098
Qy 841 RDNLTYTLVGLDKLVKP-----KNNLEAKFRQISNL-----871
Db 1099 KSNKAITSI--KVSVEPPKTKIINEIRTKSDCLAKETDLE---KQISNLSIDTQETK 1153
Qy 872 -----QELLTAFYTSAAASNWNYYQSGAKSTIIFEEIAELDPKVKKEKGVADYQVKFH 926
Db 1154 LTENGKQLKLEELLESLLKQKKNIEDQK-----KELDEVNSKIK-NIENTVNOHKCN 1205
Qy 927 YATGF-----DONAGKFN-----939
Db 1206 YEIGVINEIKIATKNQKQESTKELIKPTIQHIISSFNANDLEGIDSDENLQKINTEMG 1265
Qy 940 ---QEVIRSSSR-TIYLTSGK-----SKLEADTIDQLNQAVKNAP-LGLQSYFLDTERF 989
Db 1266 NIYEEFIKSYNLITNYLETVSKESITYNQIQNKRIDTQKELLKNVENVNAKASYLDIYE 1325
Qy 990 GVFOKATSLAVQHKQKCKTLPPKLN--NDGTYLIHDKLKKPV-----IPQISSSEPK 1041
Db 1326 NEFDRIVTH-----FKKKLATVNDNFKNKESYKVNKEGFDNISINSINTVKNSTDEN 1374
Qy 1042 WPEKLNQCSQNVNVSTFGSIIESPYFSTNFOEDADLDQGDQDSDRSQGNNSLDNQEAG 1101
Db 1375 SLNLNLQTKE-----MYANIVNNTYISYK-----EAE 1403
Qy 1102 LLKQKLAILLGNOFIOYYQOQNDKEIEFEIINVEKVSLSRVEFKLAKTL-----EDNG 1155
Db 1404 NIFRNPILKANTLNKIKNSGGIDLSKDI---KIALISY-LDSKTEDTLIFIPSPQKKT 1458
Qy 1156 KTRVLSDETWSLIVNTTIEKTPMSAVPE-----VPDTKWVEQYDPRTPLAATKPVLK 1210
Db 1459 ETVTKISDSYIILL--DILKKSQSLQKKEQQTUKLIFENRRLYEKVOATNELRGTLSDLK 1516
Qy 1211 FKQIIPVDGSGNISDKWLASIPLVIHO--QWLRLSPVVKTIRELGLKTEQOQOQOQOQOQ 1268
Db 1517 YK-----KEKILSEVKLLHKSNEKLNSCNFQNYVDITLESKSYDVQVEKSNYY 1565
Qy 1269 QQPQKKA-----VRKEEEL-ITYNPKDE-FNILNPLTKAHLRLTSLNLVN 1310
Db 1566 KQEKELIGIDPNVTDMEKEFNNDIKVIEELENNYDSSENNNLIQSKQKLKELT--NKNF 1623
Qy 1311 NDPNYKIEDLVKINEAGHQH-----APSLRANNIKRLMNTTITADYNPFP---1358
Db 1624 AEIK-KIDDKIIEKNLDLIKLETRKCNMLFTHTTLAETLK-----IKITDYSKPLESA 1676
Qy 1359 -----YNE-----DWRSDIKYLNKNGVSSHQOQAAGNQGSGL 1393
Db 1677 TKFSKEFLKYIGDTSNLSLNDIATLQKLDLHQINKVTSKLSDAT-----NDNNNL 1728
```

1394 IORLNKNIKPTPTPALIAALK-----DRNNTLSN-YSDKIIMIKPKYLVERSIGVPMST 1447
1729 IE-----KEKATOAKINLTTLFTIDSNNIDANALHNNKIOMV----- 1766
1448 GLDGYIGSOTKDGTSSSQKGFQDPTQALGLKNTYHGLGLSIRIFDPCNELAKIK 1507
1767 -----FNSLHKSIESIKOLYKMHVFLNLTGQINGKY-----FDISKQFDNLT 1812
1508 DASNNKGBEKLKLYDLFKNYLNEYKSPKIAKAGWNTNHPDKQVYPNDN-OKLPENYLN 1566
1813 Q-----LOESLNTAN-LNDLKEIGOKISDKKNKFLHALNETPIPNFNTLKEIYHD 1861
1567 LV-----LNQPKVTLNYSDDFTNLVPEPEGSDRSGTKLQVIOKQVNNYADWGSAY 1621
1862 IVKYRKQIDEIENITSEBENI--TLIID-----TITLKEKVOISILN-----F 1903
1622 LTFWYDKNIITQPNVITANIDVFTKDVKELEDNTKLIAPNITOWMPNIGSKKFKYKP 1681
1904 VTY-----ENDSNIKQHIQDNTENDVSKIESLKTTIQSFOELNKNKINGIKQAQFYDN 1957
1682 TVFFGNWENESSMNSQAQPTWEKIR-----EGFALQALKSSPDQKTRTFVLTNAP 1734
1958 ----NNININNTISTISQDVNDVKKHISKOLTIENELIEIQSLSDIKNSTVEI----- 2008
1735 LPLWKYGLPGFQNGPFTQDRLVFNQDNDNOIAALR--VQSDREKSSDKQKQWIK 1792
2009 ----RGAN-----NNYVNTIRNVEQQTNIQNNKDEIDDI 2043
1793 PKVVIPEMFNSGNIRFVGVMQIQGNTLWLPVINSVYDFYRGTDGSDNDVANLVAWP 1852
2044 QK-----ILNKNKSETKLPYI-----TGKNVNTSI----- 2070
1853 QVKTIATFNAPNVFK-----FNISKIVE 1879
2071 ----ISRKINVLNIESEYGNNNVSYNVAKLEE 2101

RESULT 5
T18372
repeat organellar protein - Plasmodium chabaudi
C:Species: Plasmodium chabaudi
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18372
R:Werner, E.B.; Taylor, W.R.; Holder, A.A.
Mol. Biochem. Parasitol. 94, 185-196, 1998
A:Title: A Plasmodium chabaudi protein contains a repetitive region with a predicted spe
A:Reference number: Z18922; MUID:98418765; PMID:9747969
A:Accession: T18372
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1939 <WER>
A:Cross-references: UNIPROT:Q25662; EMBL:U43145; MID:g1151157; PID:g1151158; PIDN:AAC634

Query Match 3.88; Score 370.5; DB 2; Length 1939;
Best Local Similarity 19.3%; Pred. No. 8.5e-08;
Matches 394; Conservative 325; Mismatches 683; Indels 643; Gaps 98;

68 SVKALIN--GKTFDPKSSFTDFVSK-----FDLTNNGRTVLEIPKKYQVVISBFS 118
4 NLKSKNGDSNKDSKKTNETSGIEKKESKNWYKIVNNTKKDKKNDSIVYDDES 63
119 PEDDKERFRIGFHLKE-KLEDGNIAGSATFIVLLPLDMPKALQGYIYVDKNFNLI 177
64 KVGENDH----HMKEVELEDQ--LKETLSITALSIVK-----EYEVKIELEKELKL 111
178 HPLNSPSAQSIKPLATRSDFI-----AKLNQFNNODELWV----- 214
112 EKEQINKNEYEK--ELNKESEFTKQRMELLKEKELNKNKNNKNNKNNKNNKNNKNN 169
215 ----YLEKQFPLEALK---ANIRIQ-----TADFSFEKGNL--VDPFYYSFIRNPQK 261
170 IESEYIEKNEKEKLYEVNTIKMSLDKLTCEVQEKDNLEKINKKVIKNNLRELKEF 229

262 ASDLNO-----DOKTVRLVLRTEFSFOAKTILDKYKYKDETFLLSSIDLKASNGTS 311
230 MKENNEIIESLDGTINDKKNAYEKLKLEISFEKRWI----- 265
312 LFANENDLKDQDLVDLLDVSYFGQSQSETITSNQVQKVPVASE-----RSLKORVFKKQOQ 368
266 -----EMLDKSLIEKENF-----ANKQAKLEKENEIIIEKLD-IESREKDF 307
369 KPRTEKPSLYEYDALSPYSOL-----OELVSKPNSIKDLVATNATLARN 410
308 KSKEEFASMENELMTLKSJLKNACQMVYKLEIKDLSQSLVEKEREIFEKNE----- 362
411 LRPSLGKTNFLFDLASHLKYVSVKAKIKOSSITTKKLFIEPLPIKISLKSSILGDQEPN 470
363 -----YDD-----KINNNKELSSINDK-----GIDNTVLHSEBK 393
471 I-KTLFEKEVTFKLDNFRDVEIEKAFGLLYPGVNBELQARQAQASFEKESKKG----- 525
394 INKLKKEKETELN-----BIHKYKNEIETIKNELNEKEB-----ELEKNNKAHTVE 440
526 ----LKEFS--QOKEENSK-----AINQOGL--BEDDNIITERLPENSPYOQENA 569
441 VTNLTKETKLEKKTEDAKEGHKNELNQLSKLNKEKONI-----KNENT 488
570 GLGASPDKPYMIKOVQORYVLAKSIOELIKAK-----DYTKLAKLJLN-RHTYNIS-- 621
489 EL-----NDKISSLNSEVNILNKDKQTLGNDIKTLNDLNNLNKNEINTSDN 534
622 --LRLEQOLFVNPRIPSPSRDIEKAFVLDKTEKNKYQIYSSASVPFQNKWSLFGYYRY 679
535 KNNKMKEDLAMLN-----EEMEGKCVVIDATEK-KY----- 564
680 LLGLDPKQTIHEL-VKLGOKAGLQEGYENLPSDFNLEDLKN-IRIKTLPFSQKNF-- 734
565 -----KNEIFMLEEKLKEK-----ENVADLNDEISI--LRNSIYVKEKFEIEMKEFYEN 611
735 KLSLDFN-----NYVDGEIKAPFGLPLFIPKELRRNSSSGGSSQNSNSPWEQEIISQF 789
612 KINLFNKPFEEKNIYENELNS-----LRLKYDN-----EOGLIKQI 648
790 KDONLSNQDLAQFSTKIWEKIIGDENEPDNNRLQYKLLKDLQESWINKTNDLY-WTY 848
649 DELNIQK-----LK-TEEKYLOLYNDNMEMPFS 675
849 LGDKLVKPKNLEAKFROISNLQELLTAFYTSAAALNNNNYQDSGAKSTIIFEB-- 904
676 ICTKIDMPYSENIKG-----SDLVDFVTAYIKRDESS--DANPDTHKEMVAELKRIA 729
905 --IAELDPKVKESKVGADVVYQLKAFHYAIGPDDNAGKFNQEV-----RSSRSTIYLTSGK 957
730 AIVAELEEKHEIA-----KLGEHGKEVVLRLGQHKKEETIILEKHK 773
958 SKLEADTIDQLNQAVKNAPLGLQSFYLDTERFGVFOKLATSLAVQHKKQKTL----- 1010
774 ----DVVTKLGEQHKENIILEEHBK-----VVTKLGDQYKEIAKLKEHAVV 819
1011 -----PKLNNNGYTLIHDKLKKPVIPOISSPEKOWF--GKLNQONGSQNVV--ST 1060
820 VAELEBKHLGEGHKEMVDELEKHAHFVEGLEKHAETAKLEEGHKSEMNEVEKHAH 879
1061 FGSIIIESPVFTNFQEDADLDODGDDSRQNNSLDNQEBAGLKKOKLAILLGNQFIYYQ 1120
880 FVEGLEEKHA-----ETAKLG--EGHREVVAGLEEKHEVVAELEBK-----HKEEIAKLE 929
1121 QNDKETEPIIINVKVSELSPRVEFK-----LAKTLEDNGKTIIRVLSDETSLIV 1170
930 EGHKEVMAEL--GEKHKEVVAGLEAKHNEEGHKEMVAELEKHAHDLAVLEEQHKAETII 987
1171 NTTIETKPEMSAVPEVDFTKWVEQYDPTPLAAKTKFVLKFKDQIPVDGSG-----NISD 1225
988 KLGEHEKVVAGIEEKYKVEATKLAEHBKDV--TKLGEQHKKEIAKLEDGHKEVNEVE 1045
1226 KWLASIPLVHQ-----QMLRLSPVVTIRELGLKTBQOQOQOQOQOQOQOQOQOQOQOQ 1281

Db 589 LINIKDEIGKNIIVFDRDRYDEVNIFVTLQLESKLQYSKWQGMDSNLKNIBSQINKTNE 648
QY 297 TFLSSIDLKASNGTSLFANE--NDLKQD-----LDVLLDVSDFYFGQ 337
Db 649 EFLSLIOIKDKGIELSESVFNDLSHIQKALDMHGSWKDELLALNKSLLDIK---VS 704
QY 338 SETITSNQVPPASERSLORVKE-----KDDQKPRIEKFSLYEYDALSPYSOLOELV 393
Db 705 SBELLSSATLK--IESLEKDVNDREMYVLTKGDIESLVIEKYK--ELKDM--YSQSDAEI 760
QY 394 SKPNSIKDLVNATLARNLAFSLGKYNFLPDDLASHLD--YYFLVSKA-----KIK 441
Db 761 L---GIKEFIN---RQTEIILKDKSVFMLELNNKFDKNNFVSKIECEDVKLKDFFKIE 813
QY 442 QSSITKKL-----FIELPIKISLKSIIIGDOEPNI-----KTLFEKVTFF 481
Db 814 SEDILANFKSDLNEFIE--SKLIQVSNIKSDNQKQIDDFLDLRISKDILNRKDSINNEVDS 871
QY 482 KLDNFRD-----VEIEKAPGLLYPG-----VNELEQARKAORASFEKES--KKGK 527
Db 872 KLSDWQSKLNEITVKIE--NLLSSGKVDLIDLSEVTTKIKELKFSIESLESYYLEKID 928
QY 528 EFSQOKEENSKATNNOGLEEDN-----ITERLPENSPIOYQENAGLGASPDUPWIKO 583
Db 929 EF-----RNOGAYSDELQIDINNHFKETRELEEN-----959
QY 584 VONORYYLAQSQELIKAKDYTKLAKLASNRHTYNISLRLEQBPDVNPRISSRDIEK 643
Db 960 -LSKFAVLNNEEFVKEVD-----SLQDKRTDIAS-----FOAN--IDITLDSLN 1004
QY 644 AKF--VLDKTEKNKYWOIYASASPVONKWSLFOYYRYLGLDPKQIHELKVLGQKAGLQ 702
Db 1005 VKPNDINKKEINGYNEVISNR-----GYSENISSKLENIEMHEIENLSRRLTDR 1054
QY 703 PEGYENLPSPFNLEDLKNIRIKTFLPS-----QKDNFKSLIDFPNYYDGEIKAPFGLPL 758
Db 1055 IDSLSK--GMDENLOKLE-----SFDVSKYQVEKPELKVKQDLD--DGEAKNK-----1100
QY 759 FLPEL-----RRNSSSGSQNSPWEQBIISQFKDONLSNOQLAQS 804
Db 1101 -LVKEIEQYKSLREBAIDYRTIDN-----DIMQAKERFGEITNELKNIESKSEFL 1152
QY 805 TKIWE-----KIIGENEPDQNNRLOYKLLDQSWINKTRDNLVYVGLDKLKVKNL 861
Db 1153 NDLYKERFKLI--ESNPEE-----RYSTFLIESEGAISKIRDEIYKLTLSN-----DENL 1200
QY 862 EAKFRQISNLQELLTAFYTSAAALNNWYQDSGAKSTIIFEEIAELDPKVE-----914
Db 1201 QIKISEMDQNFPII-----EORSKDILFEK--ELQDKIKCYCFINS 1241
QY 915 -----KVGADYVQLKPHYAIG-----FDDNAGKFNQEVIRSSRTIYLTSGSKLEA 962
Db 1242 QFGEIKAGVE--ENIKNHFDVCIKKVNLTIDDDIVKYENE-----IHKRIDSLKSIES 1292
QY 963 DTIDOLNOAVKNAPLQSGFVLDTERGVFKLATSLAVOHKQEKTLPKLNNDGYTLI 1022
Db 1293 -TFDSIE--KN-----LNDKVSGCIDKIANDFNLKYLEERCC-----NEGOJNL 1334
QY 1023 HDKL--KKPVIPQISSPEKDWPEGLKNQONQSONVNVSTFGSIIESPYFSTFOE---1076
Db 1335 ENKIDNKIKAIIDNALS--QYDGLKRYADMYDEFERLNSYIATVLSSEFKSNKEMIFEL 1393
QY 1077 -----DADIDQDQDSDRQGNNSLDNQEAGLLKOKLAILLGNQFIQYQONDKI 1126
Db 1394 ESQKLNKLNLESLDNNVKEOVIRLKEESYHNVSSHK-----LLEEDFPKDLKIRGEEL 1447
QY 1127 EFEIINV-----EKVSELSFPRVEFKLAKTLEDNCKTI-----RVLSDETWSLIVNTT 1173
Db 1448 KYSLENFIASVNDKIQNL-----EYDLSKNLENKTELICQSPRLDIEQOMKODKENFYLDFT 1503
QY 1174 IE---KTPENSAPVEFDTKWVEQYDPRTPLAAKTKFVLKFDQIPVDGSGNLSDKWLAS 1230

Db 1504 KEPSKKKDMQSEIALMETNITGKVD-----EPVDF--VNNKQSIIDSWFLN 1548
QY 1231 IPLVIHQ--QMLRLSPVVKTIIR--ELGLKTEBQOQOQOQOQOQKKAVRKEELETYNP 1287
Db 1549 IKDDVKQWQESYSTIEKRIINLAELGKTSFEND-----IFNVKIGLESF--1592
QY 1288 KDEFNILNPLTKAHLRLTSLNVLNNDPNYKIEDLKVKINEAGDHQHLAFSLRANNIKRLMNT 1347
Db 1593 KDGFEI-----KASEI--PSNL-----QNEAKKIQESVHLDKFNIGESLNL 1631
QY 1348 PI-----TFADYNPPFYNNEDWRSIDKYLNNKGNVSSHQOQAGNGQSGSL---IORLKN 1400
Db 1632 KVLDELKEVDFK-----LEKIDKVNKKTEDILIQAEVKFLTQOKDLEDKIFELNQK 1683
QY 1401 IKPETFTTALIALKORN-----NTNLSNYSDKIIMIKPKYLVVERS 1440
Db 1684 LEHEFTTLLSSNLDKVRREMVDVSSDKESFSGQIELINKNISEFSEKISLYRNN--IETS 1741
QY 1441 IGVPW---STGLDGYIG-----SQTKDGTSSSQKQFQDQFIQALGLKNTEYHGK--1489
Db 1742 IENEVNSFSKISKDLGLEDLKKSLHSTSEIETIKSGLOEQIDKEFEVFPKNHKELL 1801
QY 1490 -----LGLSIRIPDPGNE---LAKIKD-----ASNKKGEKLLKSYDL-----1524
Db 1802 KEVDNNILELESKILNCVQFNKFISEIKDNLVEYKSDLRAEFE--DSYDKINFQIENQI 1859
QY 1525 --FKYILNEYEKS-----PKIAGVNTNIHPD---OKEYPNPQKLPENYLNVL 1569
Db 1860 ENFKKLDSELEKNSIFLEAYSLKDKLEKJMETLKNEIGLQAYKKNFNVNKEFYNI--1917
QY 1570 NQPMKVTLYNSDFTITNLFPVEPGSDRGSGTKLQVIOKQVNNN-----YADMGSAYL 1622
Db 1918 ---QKETL-----GLIEIFNE-----LKEQESIKSIKNDNRRFFYSSPDSRY-1959
QY 1623 TFWYDKNIITNOPNVITANIADVFKOVKELEDNTKLIAPNITQWPNISGSKEFYKPT 1682
Db 1960 -----KSLIESYDEM-----QIYKAKIKEADEQRTILDN-----YERIS--NKESILKST 2003
QY 1683 V 1683
Db 2004 I 2004

RESULT 9
T14867
Interaptin - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T14867
R:Rivero, F.J.; Kuspa, A.; Brokamp, R.; Matzner, M.; Noegel, A.A.
J. Cell Biol. 142, 735-750, 1998
A:Title: Interaptin, an actin-binding proteint of the alpha-actinin superfamily in Dictyo
ts.
A:Reference number: Z18248; MUID:98365468; PMID:9700162
A:Accession: T14867
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1738 <RIV>
A:Cross-references: UNIPROT:O76329; EMBL:AF097019; NID:g3549260; PID:g3549261; PIDN:AAC3
C:Genetics: abpd
A:Gene: abpd
A:Introns: 173/2; 1680/1

Query Match 3.4%; Score 331.9; DB 2; Length 1738;
Best Local Similarity 17.7%; Pred. No. 3.2e-06;
Matches 331; Conservative 297; Mismatches 664; Indels 577; Gaps 65;

QY 51 DSVAPKPSIANFTSDYQSVKCALLNGKTFDPKSSSEFTDFVSKPFDFTLNNGRTVLEIPKYY 110
Db 233 DELSUTYISLFPKVVQOTLEPLNNNNISPS-----LSSSSSLLNTPNK- 278
QY 111 QVVISFSPEDDKRFRFLGFLHKEKLEDGNIQAQSATKYFYLLPLDMPKALGOYSYVDK 170


```
Qy 913 KEKVG-----DYYQLKPHYAIGFDDNAGKFNQEVIRSSRTIYLKTSQ-KSKLEAD-TID 966
Db 909 NEKNEIISVFDIY-----SDIFGGLIDEII-SNNKYLANEIDGFYVKEIDPTTG 957
Qy 967 QLANQAVKNAPLQSQFYLDTERFGVQKLAISL-----AVOH 1003
Db 958 LLSVLKK---GVYRGFYPTDRIIGIIPLLKASAKFKTTGYNLYKYVGTBYGHGHTLOQ 1014
Qy 1004 KQ-----KEKTLPLKLNNDGYTL--IHDKLKFPVPIQISSPEK-----DWFEEKLNON 1050
Db 1015 AQDVSDPSVPIPTAIGRNGISSHLHNDVLQVLDARSSGLKVRKINPDY---KPNR 1071
Qy 1051 GOSQNVNSTFGSIIESPYFSTNFQEDADLDQGDQD-----SROGN 1092
Db 1072 GVLPNYTKLLDSLKKS-----QSGKNDHGSHSNEDIWEKIDIFGSENN 1117
Qy 1093 NSLD-----NOEAGLLK-----QKLAILLGNOP-----IQ 1117
Db 1118 KTLDSNDKVLSENERFLQTVVVGIEKAAKARNLQYDLFILNSFDHESATINPGISGKE 1177
Qy 1118 YVQONDKEIPEIINVEKVSELSFVFEKLAKTLEDN-GKTRV-----LS 1162
Db 1178 YFKLDK---VEFFPAEDENSTDIKSFPEKTDVLSNGLNVKFDKDKRIIATYKEDLA 1234
Qy 1163 DETMS-----LIVNTTIEKTPMSAVPEVFDTKWVEQYDPTPLAAKTFV- 1208
Db 1235 NDTFSDLKPTTPIAKDKPLDLSIFEKTKSLKALKKEIEEK-----QRLLENEPEN 1285
Qy 1209 LKFKQIIPVDG-SGNISDKWLASIPLVHQMLRLSPVVKTRIRGLTKTEQOQQOQQOQ 1267
Db 1286 LDFQTTFSVNGWSGLDKKLTKIB-VIHEE-----NNELNENHLODFK 1328
Qy 1268 QOQPOKKAVRKE-----BELETYNPKDEFNIIPLTKAHLTLNGLNVNDN-N 1314
Db 1329 EFVETGRIKKYVDFTKLENFYQYLRKHGKANFTF-----YQLLSTLITGQKVT 1381
Qy 1315 YKIEDLVKVNAGDHQALFSL-----RANNIKELMNTPTIT-FADYNPFYFYNED----- 1363
Db 1382 LKGKEYKLTNDQDELWLRAKILHDPSRFGIDELIKORTSYYPNYSIFHFENADYFVN 1441
Qy 1364 -----WRSIDKVLANKGVSSHQOQAGNGOGSLIQRLNK----- 1399
Db 1442 PRKENRLKEQIIRGKKNYSLDIALSNYKONSHENE-----NKKATHY 1487
Qy 1400 -----NIKPETFTPALIALKDRNNTNLSYSDKIIMIKPKYLVERSIGVPWSTGLDG 1451
Db 1488 YNGEFNEFKVKGIDGKEPIILDADWYEHFNFGDNHFKENFFRTK--ENESI-----TSFDS 1540
Qy 1452 YIGSEQ-----TKDGTSSSSQO-----KGFDDQDFIOALGLKNTE 1485
Db 1541 FESPFESSIDPLKFTLTGQGENAQRWDIDYALTKFNLDYAKKLDSDLFKTTLDQSK 1600
Qy 1486 YHGKLGLSIRIFDPGNELAKIKDANKKEGKLLKSYDLFKNYLNEYEKSPKI----- 1539
Db 1601 KYDKAAL-IOAPE-----KAYGQEKAKEL---QDIANDLMDKFESSLAFPMKNHN 1648
Qy 1540 -----AKGWTNIHPQKEVPNOKLPENLYNLVNLQNPWKVTLNYSDDFI 1584
Db 1649 IKDLSVDHLFVSOVGILGFNNITKSLR--PNPNDN-----NWVIER-WKTSPISSDVI 1699
Qy 1585 TNLFVEPEGRSGSTKLGKQVTKQVNNYADWGSAYLTTFWYDKNIIINQPNVITA-NIA 1643
Db 1700 LAWKTDNPSND---SRIVMDAIRKFLKDNNDID--DSKLDLFLKYLILGTQ---ILNGKND 1752
Qy 1644 DYFIKD---VKELED-NTKLIAPNITOWMP-----NISGKSKFYKPTVF----- 1684
Db 1753 QIFIENNFOINELNKNISRFENDFGQYFSDVFNFAESLRTDYVQVTSYVPSQOLKMN 1812
Qy 1685 -----FGWENENSMNSQAQPTTWKIREGFALQALKSSFDQKTRTFVLTNAPLPLWK 1739
Db 1813 GYLKGFSEWNTGNEYVVSGLTKW-----LNQINVLSD-----LLRRRSIYSRWS 1859
```

```
Qy 1740 YGP-----LQFGQPNFKTQDWRLVFNQDNDNTAALRVQE-----Q 1775
Db 1860 YVEQASQFLHFGNAQVFNYSKKL-----DEVLSLTTKKBFKFDQSQSPLVNYFELLSYQ 1914
Qy 1776 DRPEKSEDKDKQKWKFKVVIPEEMFNSGNIRFV 1810
Db 1915 EKQKLNSESTK-----TEELDRSDLKYYI 1940

RESULT 13
S38173
myosin-like protein MLP1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YKR095w; protein YKR415
C:Species: Saccharomyces cerevisiae
C:Date: 03-May-1994 #sequence revision 03-May-1994 #text_change 09-Jul-2004
C:Accession: S38173; S40647; S31207
R:Baladron, V.; Balalesta, J.P.G.; Bou, G.; del Rey, F.; Eteban, P.F.; Garcia-Cantalejo,
submitted to the Protein Sequence Database, March 1994
A:Reference number: S38158
A:Accession: S38173
A:Molecule type: DNA
A:Residues: 1-1875 <BAL>
A:Cross-references: UNIPROT:Q02455; EMBL:Z28320; NID:9486586; PID:9486587; MIPS:YKR095w
A:Experimental source: strain S288C
R:Bou, G.; Eteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; Jin
Yeast 9, 1349-1354, 1993
A:Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chromo
A:Reference number: S40644; MUID:94205265; PMID:8154186
A:Accession: S40647
A:Molecule type: DNA
A:Residues: 1-1875 <BOU>
A:Cross-references: EMBL:X73541; NID:9450550; PIDN:CAA51948.1; PID:9450554
A:Experimental source: strain S288C
R:Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A:Title: A new yeast gene with a myosin-like heptad repeat structure.
A:Reference number: S31207; MUID:93247549; PMID:8483450
A:Accession: S31207
A:Molecule type: DNA
A:Residues: 1-300, 'A', 302-1875 <KOE>
A:Cross-references: EMBL:L01992; NID:g171958; PIDN:AAA34783.1; PID:g171959
C:Genetics:
A:Gene: SGD:MLP1
A:Cross-references: SGD:S0001803; MIPS:YKR095w
A:Map position: 11r

Query Match 3.1%; Score 305.5; DB 2; Length 1875;
Best Local Similarity 19.6%; Pred. No. 4.5e-05;
Matches 340; Conservative 271; Mismatches 627; Indels 494; Gaps 75;

Qy 60 ANFTSDYQSVKKALLNGK-----TF-----DPKSSEFTDFVSKFPL-----TNNGR 101
Db 380 AKSSDSIFLKLQIKERRTKHEHLQNIQTFIVELEHKVPIINSFKERTDMLNELNAA 439
Qy 102 TVLEIPKQYVIVSEPSDEDDKERRFLGRFLHKEKLEGNIAQSAATKIYLLPLDMPKAA 161
Db 440 LLEL-----HTSEKNNAKVKELNAKNQKVECNLDQLTLTK-----QRLDLCR--- 482
Qy 162 GOYSYIVDKNFNNLIHPLNFSQAQSIKPLALTRSSDFIAKLNO-----FNNQDELWVYL 216
Db 483 -QIQYL-----LITNSVNSDSKGPLR---KEEIQFIQNIQEDDSTITESDSQKVV 530
Qy 217 EKFFDLBALKANIRLOTADFSFEK--GNLVDPFVYSFIRNPQKQEWASDLNQDQKTVRL 274
Db 531 ERLVEFKNI---IQLEKNAELLKVVVRNLADKLEKSKQSLQKTESVTVNEAKEAI 587
Qy 275 YLRTEFSPQAKTILKDYKYDE--TFSSIDLKASNT--SLPANENDLKQDQVLDLVD 330
Db 588 TLKSEKVDLSRTEELQKLEELKTSVPNEADASYNVITIKQLTETKRDLESQVQDLQTRI 647
Qy 331 SDYFGGQSEIT--SNSQVKVPASERSLKDRVFKQDQKQPRIEKFSLEYDALSFYSQL 389
Db 648 SQITRESTENSLNKEIQDLYDSKDISIKLGKSKSRILAEERFKL----- 695
```



```
QY 320 -----KDQDVLDD-----VSDYFGQSETIT--SNSQVFPVPASERS 356
Db 1059 LBSVIAEKEQLKTDLKENIEMTIENQEBELRLGLDELKQOEIVAOEKNHAIKEGELSRT 1118
QY 357 LKDRVKFKDQOKPRIEKSLSVEYDALSFYSOLOELVSKPNSIKDILVNATLARNLRFSLG 416
Db 1119 C-DRLAEVEKLEKESQQLQOEKQOQLLVQEBEMSEMKKINELINKB--LKNKELTLE 1175
QY 417 KYNFLPDDDLASHLDYFLVSKAKIKOSSITKKLFIELPIKISLKSILGDQEPNITKTFE 476
Db 1176 HMEITERLELAQLNENYBEVKSITTKERKVKEL-----QKSEF 1213
QY 477 KEVTFKLDNFRDVEIEKAGLLYPGVNHELEQAKAQRASFEKESKKGLKFSQOKEEN 536
Db 1214 TERDHLRGVIREIE--ATGL-----QTKBELKTAH-----IHLKEHOETIDELRSVSEK 1261
QY 537 SKAINNOEGLEEDNITERLPENSPFOYOENAGLGASPDKPMIKDVQONQRYLAKSOI 596
Db 1262 TAQIINTQDLEKSHT--KLOEIPVLHBEQEL-----LPNVKVSQETQETMNELELL 1311
QY 597 QELIKARDYTKLAKLNSRHTYNISLRLKEQFDVNPRIPS--SRDIEKAKFVLDKTERNK 655
Db 1312 TEQSTTKDSTTLARIEMER-----LRLNEKFOESQOEIKSLTKERDNLTKIKEALEV-K 1364
QY 656 YMOIYSSASPVFQNKSLFGYRYLLGLDPKQTIHE-LVKLGQKAGLQFEGYENLPDSE 714
Db 1365 HDQL-----KEHIRETLAKIQESQSKQ-----EQSLN 1391
QY 715 LEDLKNIRIKTLPFSOKDNFKLSLDFNNYDGEIKAPFGLPLFLPKELRNRSSNGGS 774
Db 1392 MKSKDN--ETTKIVSEWEQFKPK-----DSALLRIELEM-LGJSKRLQESH-----1435
QY 775 QNSNSPWEQBIISQFDQNLNSQDLAQFSTKIWE---KITDENEFQNNRLQYKLLKD 831
Db 1436 -----EMKSVAKEKDDQLRLQEVQLQESDQLKENIKEIVAKHLETEBELKVAHCLKE 1488
QY 832 LQESWINKTRDNL-----YWTYLGDKLVKPKNNLEAK-----PROISNLQELL 875
Db 1489 -QBETINELRVNLSEKETEISTIQKLEAINDKLQNKIQBIYEKBERQLNIKQISEVOENV 1547
QY 876 TAF-----YTSALSNNNVYQDSGAKSTIIFEBIEAELDPKVKKEKGVADV 920
Db 1548 NELQKQEKHRKAKDSALQSTESKMLELTNRLQESQEBIQIMIKEEM-KRQVEALQIER 1606
QY 921 YQLKFHYAIGFDDNAGKFNQEVIRSSRTI-YLKTSKSKLEAD--TIDOLNQAQVNAPL 977
Db 1607 DQLK-----ENTKEIVAKMESQESQEQYQFKMTAVNETQSKMCEIEHLKEQFETQKL 1658
QY 978 GLQSFYLDTERFQVFKLATSIAVOH-----KQKEKTLPKLNNQGYTLIHDKLKK 1028
Db 1659 NLEN--IETENIRLQTLHENLEMRSVTKERDRLRSVEETL--KVERD--QLKENLRE 1711
QY 1029 PVIPOISSPEKDFEGLKQNGOSQNVNVSTFGSIIESPYFTNFOEDADLDQDQ---1084
Db 1712 TITRDLKQEBELKIVHMLKHEH--QETIDKLRGIVSEKTNELSNQKXOLEHSDALKAQ 1768
QY 1085 ----QDSSROGNSLNDQE-----AGLLKQKLAILL-----GNQFTQYYQONDKEI 1126
Db 1769 DLKIQEBELRIAHMLKEQOETIDKLRGIVSEKTDKLSNMQKOLENSNAKLQEIQLKAN 1828
QY 1127 EPELINV-----EKVSELSFRVEFKLAKTLEDNGKTTIRVLSDETMSLIVNTTIKTP 1178
Db 1829 EHQLITLKQDVNETQKQVSEME-----OLKQKIQOQSLTSLKSLEIENLNL-AQELHENLE 1882
QY 1179 EMSAVPEVDT-KWVEQ-----YDPRTPLAAKTKFVLKFKDQDQIPVDG 1219
Db 1883 EMKSVMERDNLNRVBEETLKERDQKESQETKARDLEIQOELKTARMLSKHEKETVD-1941
QY 1220 SGNISDKWLASIPLIVHQMLRLSPVVKTIRELGLKTEQOQQOQQOQQOQQOQKAVRKE 1279
Db 1942 --KLREK-----ISEKTIQISDIQK-----DLDSKDBELQKKIQELQKQELQLLRVK 1986
QY 1280 BELETYPKDBFNILNPLTKAHLRTLSNLVNDPNYKIBDLKVINKBAGDHQLAFLSRAN 1339
```

```
Db 1987 EDVNMSHKK--INEMEQLKQOF-----EPNY-----LKCENDMNFQLTCKLH--2026
QY 1340 NIKRLMNTPIITFADYNPPFYNNEDWRSIDKYLNNKGN--VSSHQOQAAAGNGGSLIORL 1397
Db 2027 --ESLEEIRIVAKE-----RDELRLRIKESGLKMERDQFIATLREMIARDRO-----2069
QY 1398 NKNIKPEFTFALIALKDRNNTNLSYSDKIMIKPKYKLVERSIGVPWSTGLDGYIGSEQ 1457
Db 2070 NHQVPEK-----RLSDGQOHLMESLREKCSRIKE--LLKR-----YSEMDDHV--2112
QY 1458 TKDGTSSSSQOKGPDQFIQALGL---KNTEYHG-----KLGLSIRIFDPGNELAKIKDA 1509
Db 2113 -----ECLNRLSLDLEKEIRHFRIMKKLVLSY-----VTKIKEE 2148
QY 1510 SNKKGEEKLLSYDLFKVNLNBEYEKSPKIAGTWNIIHPDQKEYPNPNQKLPENYLNVL 1569
Db 2149 -----QHECINKFEM--DFIDEVEKQKELLK---IQHLQOQCDVPSPRELAD---LKL 2193
QY 1570 NQPKVVTL-----YNSDF-----ITN-----LFVEPEGSDRGSGTKLKQVIOK 1608
Db 2194 NQMDLHIEELKDFSESEFPISIKTEFOQVULSNRKEMTQFLSEWLNTFRDIEKLKNGIOK 2253
QY 1609 -----QVNNNYADWGSAYLTFWYDKNI--ITNQPNVI---TANIADVFIVKVKEL-EDMT 1657
Db 2254 ENDRICQVNN-----FFNNRIIAIMNESFEERSATISKWEQDLKSLKEKNE 2302
QY 1658 KLIANITOMWPNISGSEKFKYKPTVFFGNWENENSNNSQAOQTWTKIRE 1709
Db 2303 KLFKNYQTLKTSLASGAQ---VNPT---TODNKNPHVTSRATQLTTEKIRE 2347
```

RESULT 16

S05603

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (str S05603)
N:Alternate names: gp195 surface antigen
C:Species: Plasmodium falciparum
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
C:Accession: S05603; S04850
R:Wyller, P.J.
submitted to the EMBL Data Library, April 1989
A:Reference number: S05603
A:Accession: S05603
A:Molecule type: mRNA
A:Residues: 1-1639 <MYL>
A:Cross-references: UNIPROT:P04933; EMBL:X15063; NID:G9896; PIDN:CAA33163.1; PID:G9897
R:Wyller, P.J.

Nucleic Acids Res. 17, 5401, 1989
A:Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Plasmodium falciparum.
A:Reference number: S04850; MUID:89345116; PMID:2668887

A:Molecule type: mRNA

A:Residues: 1504-1639 <MYL2>

A:Cross-references: EMBL:X15063

C:Superfamily: major merozoite surface antigen

C:Keywords: glycoprotein; merozoite; surface antigen

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-1639/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 3.0%; Score 295; DB 2; Length 1639;
Best Local Similarity 19.5%; Pred. No. 0.0001;
Matches 282; Conservative 204; Mismatches 478; Indels 484; Gaps 68;

QY 203 LNQFNQDELWVLEKFFOLEALKANIRLOTDAPSEKGNLVDPPVYFIRPNQKEWA 262
Db 168 IDGYEBINELLYKLNFFDLRAKLN-----DVCANDYQCFPNLJKIRA 211
QY 263 SDLNQDQKTVRLVLRTEFPQAKTILKDYKQDETFLSSIDLKASNGTSLFANENDLKQ 322
Db 212 NELDLVLUKLV-----FGYRKPL-----DNIKDN 234
QY 323 LDVDDLDSVDYFGQSETITNSQVFPVPASERSLKDVRKFKDQOKPRIEK-----FSLY 378

QY 482 -----KLDNFRDVEIEKAFGLLYPGVWNEELEQARKQAFSEFEKSKK 524
Db 480 SRTQCSLSREVICLMAELHDLNETKSRNPATVQVALDEYAQNPFSTASSETLVNKE----- 535
QY 525 GLKFSQOKEENSKAINQOE---GLEEDDNITERLPENSPIOYQOENAGLGASPKPYMI 581
Db 536 -LANFSSIKAEAVSKTLEURKVRALCEVIEI-----OKOTVQYQISNAVKENSNTLSQOI 589
QY 582 KDQVQRYVYLAQSQI-----QELIKAKQYTKLAKLLS-NRHTYNISLRLEKQELFDVN 632
Db 590 KMLESE---LNSSKIKNESLLNERNLLKEMLATSRSSILSHNSSAGNIDDKMS--ID-- 642
QY 633 PRIPSRDIEKAKFVLDKTEKNKYQIYSSASPVFNKWSLF-----GYRYLLGL 683
Db 643 ---ESTRELEKNYEVY---RNEMTAIOESLSKRNQDLSLSEMAIRKELENSKYOOQLST 695
QY 684 DPQOTIHELVLGOKAGLQFEGVENLPDNLEDLKNIRIKTPLPFSQD-----NFKLSLL 739
Db 696 DRLTNANDVEAFK-----EAKELRSINQNLQD-----IISROQARASKFABELL 741
QY 740 DFNYYD---GBIKAPEFGLPLPLKELRRNSNGSQNSQNSPWEQE-IISQFKD-QNL 794
Db 742 HVNSLAERLKGELNASKG-----EKDLAKRTQERLISENDKLLAERERLMSLVSDLOTF 795
QY 795 SNQDLAQFSTKIWEKIIQDENEF-----DQNNRLO---YKLLKDIQESWINKTRDN 843
Db 796 LNOOQLSDAARKV--KFSESEKSLSLQKLKESNEKMSNDLHLSLQKSLKESGIEYS--S 851
QY 844 LWTYLGDKLVKPKNN-----LEAKFROI-SNLQEL-----LTAFTYSAAALSNNW 888
Db 852 RIKTLMLEKQSEDRNKLNDQOMMEIKLOELGVIELEKORFSTLEAKFTQ---OKQT 908
QY 889 NYQDSKAKSTIIFEBIAELDPKVEKVGADVVQLAFHYAIGFDDN---AGKFNQBVIRS 945
Db 909 SYSE-----REALLESSLDLSQKHTS-----LESQYNTSLRNIQLOQAASKLAEMV-E 957
QY 946 SSRTIY---LQTS---GSKLEADIDQ-----LNOAVKNAPLGLQSFYLDTERFQVFK 994
Db 958 RVKTEYDEVRLQTSLEKHNHAKITSLEQRVILQDEIASSLRNCENITKDSB----- 1010
QY 995 LATSLAV-----OHKQEKTLPKKLNNNDGYTLIHLKDKKPVIPQI 1034
Db 1011 --TRVALLLEENKHNELSSHRNAKHELENDYKQOL-----LLVTEDLRK----- 1057
QY 1035 SSSPEKDMFEGKLNQSQSNVNVSTFGSIIESPYFTNFQEDADLDQD-----GQDSSR 1089
Db 1058 ---TREDYEKELLRHADAR---STLQKLRED--YTKALEQVEDLNKETALKAGINESQ 1107
QY 1090 ---QGNNSLDNOEAGLLKQKLAILLGNQFIQYYQONDKEI----- 1126
Db 1108 PFPPISEKEDPL-RQEVVYLVKQONAMLLTQ--LOSSNLNFAEITSPSPDLDSVMKGLGSLD 1164
QY 1127 -----PFEIINVEKVSLEFRVEFKLAKTLEDNGKTIIRVLSDBTWSLI VNTTIKT 1177
Db 1165 QNHVRIKSEMEIISQOR--QLLFLENKLRIVESSNRVIADLQRGITEKDVSSSTSESV 1222
QY 1178 PEMSAPVEVDTKWVEQYDPRTPLAAKTFVLKFKQIIPVDSGNSISDKWLASIPVTHQ 1237
Db 1223 GERS-----NYLNMVALLNESNKS-----LRE 1244
QY 1238 QMLRLSPVVTIREL--GLKTEQ---OQOQOQOQOQOQOQKAVRK-----EBELEYNPK 1288
Db 1245 NLERNNEVITELREKIEITLTDLANFRLNKEQLESQLOTEKAAYVKLENSNEEYKGRHOE 1304
QY 1289 -----DEFILNPLTKAHLRTLSLVNNDPNYKIEIDLKVNKEAGDHOIAFS 1335
Db 1305 ILLSLNSSTSSDASKLNELVSKENLI--EELNQEIHLKSELETVKSKSEDL---N 1359
QY 1336 LRANNIKR-----LMNTPTITFADYNPFYFYNNEDWRSIDKYLNNKGNVSSHQQAAGNQG 1390
Db 1360 ERAQNSQKIEOLEKNTKLAIAA-----WRTKYEQVWVK-SLEKHNQIRQOOLSQK 1407
QY 1391 SGLIQ-----RLNKNKIKETFTFPALIALKDRNNTLSNYSDDKIIMIKPKYLVERSIGV 1443

Db 1408 TSELEAKVAECHQNEQLNKFSATPTATTQSEPTSVSLEEFN----- 1449
QY 1444 PMWSTGLDGYIGSEQTKGTSSSQOQPDQDFIOALGLKNTYHGKLGSLSTRIPDGNEL 1503
Db 1450 -----STKEELSSQTKQ-----LSEIMDILNT-----TKREL 1476
QY 1504 AKIKDASHKK-----GEEKLLKSYDLFKNYL-----NEVEKKSPTIAK 1541
Db 1477 EKVQRNSNKSEGTSKDTIPNEEEMERKVKVQEQEVLRLSRIAKELQKVELLRKQNVLO 1536
QY 1542 GW-----TNIHPOQKEYPNPNQKLPENYLNVLNQPKWVLTLYNS--SDF 1583
Db 1537 DQVKALQETVVSSEEAESASVHADTKDL--BNLKKTTEMLSVTFQ-----VIFNESISDF 1589
QY 1584 ITN-----LFVEPEGSDRGSGTKLKQVIOKQVNNYADWGSAYLTFWYDKNIIITQPNVI 1638
Db 1590 STSTADFTTFVQKEWEKR-----REILQKVVEQVAQSHQKQLD----- 1628
QY 1639 TANIADVFIDKVELEDNTKL-----IAPNITQWPNENISGSKEKFKYKPTVFFCGNENESS 1694
Db 1629 --NTR-----KELEMNKLKLSMLEKNKLARVRAELQSKKK--DSPAILLSLEASKNTDS 1678
QY 1695 MNSOAOQTPTWEKIREGFALQALKSSFDQKTRTFVLTINAP 1734
Db 1679 NKSNEVP-----AAQVKEKKLIAKTHS--VDTNSP 1707

RESULT 18
A24594
Probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodium f
C:Species: Plasmodium falciparum
C:Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jul-2004
C:Accession: A24594
R:Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholls, S
Nature 317, 270-273, 1985
A:Title: Primary structure of the precursor to the three major surface antigens of Plasmo
A:Reference number: A24594; MUID:86014355; PMID:2995820
A:Accession: A24594
A:Molecule type: DNA
A:Residues: 1-1640 <HOL>
A:Cross-references: UNIPROT:P04933
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 3.0%; Score 291; DB 2; Length 1640;
Best Local Similarity 19.2%; Pred. No. 0.00015;
Matches 322; Conservative 221; Mismatches 543; Indels 592; Gaps 79;

QY 203 LNOFNNQDELWVLEKFFDLLEALKANIRLOTADFSPEKGNLVDPFVYSFIRNPONOKEWA 262
Db 168 IDGYEBEINELLYKLFYFDLLRAKLN-----DVCANDYCOIIPFNKIRA 211
QY 263 SDLNODQKTVRLYLRTFESFQAKTILKQYKDETFDLSSIDLKASNGTSLFANENDLKQ 322
Db 212 NELDLVLKLV-----FGYRKPL-----DNIKDN 234
QY 323 LDVVDLDVSDYFGGQSETTSSQVQKVPAPGSRSLKDRVKFKDQOKPRIEK-----FSLY 378
Db 235 VG--KMEDYIKONKKTIENTINEL--IEB8KTTDKKNKATKEEKKLYQAOYDLSY 288
QY 379 -----EYDALSFYSOLOSLVSKPNSIKDLVNAT----- 406
Db 289 NKOLEBAHNLSVLEKRIOTLKKENIKELLDKINEIKNPPPPANSNGTNTLLDKNKKIE 348
QY 407 -----IARNLRFSLGKYNFLFDLASHLDYFYLVSKAKIKQSSITKKLFTIELPIKIS 458
Db 349 EHEKEIKETAKTIKFI---DSLFTD-PLELEYL--REKKNV-----IDISAKVE 393
QY 459 LKSSILGDOQEPNIKTULFEKVEVTFKLDNFRDVEIEKAFGLLYPGVNEELSQARK---AQR 514
Db 394 TKESTBPNEYPN-----GVTYPLS-----YNDINNALNELNSFGDLINP 432

[illegible][illegible]

A;Molecule type: DNA
A;Residues: 1-1979 <GAR>
A;Cross-references: UNIPROT:O96133; GB:AE001375; GB:AE001362; NID:G3845105; PIDN:AAC71819
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0145C

Query Match 3.0%; Score 290.5; DB 2; Length 1979;
Best Local Similarity 17.1%; Pred. No. 0.00021;
Matches 293; Conservative 301; Mismatches 559; Indels 563; Gaps 73;

Qy	248	VYSFIRPNQKQEWASDLNQDQTVRLYLRTSPQAKTILKDYKYKDETFLSSSIDLKAS	307
Db	2	VFTF-KNKKKKKQSSD-----KVSKSFNEEDNENNEKKREKSDSYWKLIETKKG	51
Qy	308	NGTSLFANENDLKQDLQDVLDDVSDYFGGQSETITNSQVQVPASERSLKDRVKFKQ	367
Db	52	SKTK-YKDNLSLDNINEDIINNNDNNNDNNNDNN-----NNDNNNDNNNDNNE	104
Qy	368	QKPRIKFSLYEYDALSFYQOLQELVSKPNSIKDLVNATL--ARNLRFSLGKYNFLPDDL	425
Db	105	NNNDNNNNFNNYS-DEIG-----KNIIHKNELENLQKDTLKSSISLGNKINVNYESKIEEL	158
Qy	426	ASHL-----DYFLVSKAKIKOSSITKKLFIELPKISLKSIIIGDQEPNI---	471
Db	159	EKELKEYKDKNIDNNDY---ENKLKEKEDFVKQKIDM-----LNEKENLLQEKELDINKR	210
Qy	472	-KTLFEKE--VTFKLDNFRDVEIEKAFGLLYPGVNEELEQAKQAQASFEKESKGLK-	527
Db	211	EKKINEKKNIIKKEEFTNIE-----KEYLEKNKERETISIEIIDIKKHLEK	258
Qy	528	---EFSQOKES---NSKAINNOGLEE-----DDNITERLPENSPIOYQ	565
Db	259	LKIBIEKKEDLENLKKLSKENVNLKELGCVKKEKNETINSLNDNIIKEKKYKLLEYE	318
Qy	566	QENAGLAS-PDKPYMIKDQVQORYYLAQSOIELIKAKDYTKLAKLSNPHYTNISURL	624
Db	319	LEENKQKIDLLNKQEKKEKEKER---EKEKEREKEKEKYDTLTIKSKOB-----KISI	370
Qy	625	KEQLFDVNPRISSRDIEKAFVLDKTEKQNMWQIYSSASPVFQNKWSLFGYYRYLLGLD	684
Db	371	LEKVSHTKVR---EMWIEKRE-----	388
Qy	685	PKQTIHELVLKQKAGLQFEGYENLPDSFNLEDLKNIRIKTLPFSQKDNFKLSLLDFNNY	744
Db	389	-----HNFLHMDQLKDLKNSFV-----KNNNLKV-----	414
Qy	745	YDGEIKAPEGLPLFLPKELRRNSNGSGSONSPWEQEIISQPKQNLNSQDOLAQFS	804
Db	415	YKCBIKN-----LKTEBKKEKELKDIEVNSKEEINKLINQLEK-----EKQILAFN	462
Qy	805	TKIWEKLIIGDENFDQNNRLQYKLLKDLQESWINKTRDNLNWTYVLGDKLVKPKNNLEAK	864
Db	463	KNHKEEIHGLKEELKESVKTIETQELQEM-----	497
Qy	865	PROISNLQELLTAFYTSAAAL-----SNNNWYQDSGAKSTIIFBEIAELDPKVKKEVGAD	919
Db	498	QKELDQLEKYNQAQIESISIELSKKEYNQY-----KNTYI-BEINNLEKL-BETNKE	550
Qy	920	VYQLKPHYAIGFDDNAGKFNQEVIRSSRTIYLTSGSKSKLEADTIDOLQAVNKNAPLGL	979
Db	551	YTNLQNNY-----TNEINMLNND-IHMLNGNIKTMTQI	583
Qy	980	QSFYLDTERFGVQKLATSLAVQHKQEKTLPKKLANNDGYTLIHDKLAKPVIPOISSPSE	1039
Db	584	STLKNDV-----HLNLEQI---DKLNNKGTLL-NSKISBLNVQIMDLKEE	624
Qy	1040	KDWFBGKL-----NQNGSQNVNVSTFGSIIIESPYSTNFQE	1076
Db	625	KDFLNNQIVDLSNQIIDLITRWEEKENQMLQENKYQOEMELLR-GNIKSSENILANDEE	683
Qy	1077	DADLDQDQDDPSRQGNNSLDNQEAGLLKQ-----KLAILLGNQFOYQOONKEIEFEIIN	1132

Db 2950 ETDNSGGVNLKFKKDSBSTKRN-----FISNLTITKQNLNV 2991

RESULT 24
F90559
conserved hypothetical protein MYPV_3820 [imported] - Mycoplasma pulmonis (strain UAB CT)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: F90559
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galieson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: F90559
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1183 <KUR>
A:Cross-references: UNIPROT:Q98QH9; GB:AL445566; PID:g14089796; PIDN:CAC13555.1; GSPDB:C
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPV_3820
A:Genetic code: SGC3

Query Match 2.9%; Score 287; DB 2; Length 1183;
Best Local Similarity 19.8%; Pred. No. 0.00014;
Matches 308; Conservative 217; Mismatches 482; Indels 552; Gaps 71;

QY 71 KALLNGKTFDKSSEFTDFVSKDFLTNGTNGTVLEIPKQYV-----ISESPEDDK 123
Db 4 KKTIASSFLPMSLLF-----FLSNTSTIAQEKNIQVFKKNSLENDFEYRKND 55

QY 124 ERFRGLFKLEKLEGNIAQATKFIYLLPLDMPKAAQGYSIYVDKFNFNLIHPLSNF 183
Db 56 E-----IGEK-----TLSIVSKFIETKLNLYKSDFEKINYIVINDSN-----ENF 97

QY 184 SAQSIKPLALTRSSDFIAKNOFNQDELWYLE-KFPDLBALKANRLQTADESF-EKG 241
Db 98 S-----LGNNSQIENWTNLNVKNTDID-----KLSNVKFFFNKDS 133

QY 242 NLVDPVYFIRNPQNKQEWASDLNODQTVRLVLRTEFSPQAKTILKDYKYKDETFLSS 301
Db 134 SOSWDFIEETINIKORIEKISLDSQ-----KYNNDHFSVS 169

QY 302 -----IDLKASNGTSLFANENDLKQDLVDLDVDFGGQSETITTSNSQVFPVASE 354
Db 170 IKKLDPIYDYSKPNDSLFLNE--PRKKVDV-ILEIKENVYKKEI---NSQIQ----- 217

QY 355 RSLKDRVFKKDDQKPRIEKFSLEYDALSQYQLQELVSKPNSIKDLVNATLARNRFS 414
Db 218 -----KIVKDFDK--LNFETSSLLFWKENVNLDDLHNEYIKKIQAIDFNKPKFE 267

QY 415 LGKYNFLDLDLASHLDYFL-----VSKAKIKOSSITTKLFTLPIKI 457
Db 268 I-KVQILSDDLISS-DVFLLLFRGEYFNQVNIIDLGRVKIKSRSDINSQSPIQRILLI 325

QY 458 SLKSSILGDQBNIKTLFEKFTKLDNFRDVEIKAFGLLYPG-----VNELEQARKAQR 514
Db 326 YKKNVYKNFDG--KNGFEKVPQPIHEEREISNE-----YGGHWLVNQ-----PPR 370

QY 515 ASFEKEKSKGLKEFSQKEB-----NSKAINNQGLEEDDNITERLPNSPIQOENAG 570
Db 371 MAFTS-----SDQNEVIYVDDIAIDNIDNLFPPD-----LEHQQRKKK 409

QY 571 LGASDPKPYMKDVQNRQYLAQSIQELIKAKDYTKLA-----KLLSN 614
Db 410 TTKITIK--TKDTNKTLY-----EIIYSKELKDSAIKARFNWDPPINLHKKLVD 458

QY 615 RHTYNISLRLEQLFDVNPRISSRDIEKAKFLVDKTEKNKYQWYSSASPVQNKWSLF 674
Db 459 KHNENGVNPK-----YNPRINPHNGLEBKIFWLEHEKLN-----HLSNPKKVPF-- 504

QY 675 GYRYLGLDLPKQTHLVKLKGQA-----GLQFEGY-ENLP-SDFNLEDLNKR 722

Db 505 -----NIKEGAFVKVSGISQKAYSIDLPPENSLTKKYFWNLNNDFSQBELITIK 555

QY 723 IKTPLPFSQKDNFKLSLLD---FNNY-----YDGEIKAPERGLFLFPKPE 763

Db 556 DNNNSFSKSGNFFLELKNYFNYSNLIISVGSNKLHNLDELIDKKILVP-----VEKS 608

QY 764 LRRN-----SSNSGGSQNSPWEQEIISQFKDQNLNQDQLAQFSTKIWEKIIGBENE 817

Db 609 IAGNILKNYLLKNFPESISSLSYEIVQNVKLLNVLNLEFLNKNKYNKKKIFDIENK 668

QY 818 FQNNRLQYKLLKDLQBSWINKTRDNLWYTLGDKLVKPKNNLEAKPRQISNQLLELTA 877

Db 669 MDSLS-----TVDNKNKL-----ISLIDE--EE 690

QY 878 FYTSAALSNWNYQDSGAKSTIIFEBIAELBPVKKEKVGADVYQLKF-HYAIGFDDNAG 936

Db 691 FLKINSFKNTKNF-----VEID-----YQDHPNVDLTFEKKK- 723

QY 937 KFNOEIVRSSRSTIYLTSGKSKLEADTIDQLNQAVKNAPLGLQSFYLDTERFGVFQKLA 996

Db 724 IFYKDKINKKQITI-----NINKERIIEINSLIKDG----- 754

QY 997 TSLAVQHKQKEKTLPKLNNNDGYTLIHDKLKVPVIP-QIS-SSPEKDWF----- 1043

Db 755 -----KKINSNIANLIPDSTDKKIDILKINISRKEDLYLVEYLFPHVKDY 799

QY 1044 -----EGKLNQSQSONVSTFGSIIESPYFSTNFOEDADLDDQD 1084

Db 800 QDYPYLSYDVYHKFILEDNKNSFLSSKKNFNOQIENII-----FLEIADLNLNG 850

QY 1085 QDPSROGNSLDNQEAGLLKQKLAAILGNFIOYQON-----DKEIFEIINVEKV- 1136

Db 851 ETNFEK-----IKNQILSHLEKQVANGSTNFDKE-KPLVLNNENLK 889

QY 1137 -----SELSPRVFEKLTLEDNKTIRVLSDETMASLIWNTTIETKTEMSAVPEVD 1188

Db 890 LANKVFLNKEIPEKIETL-----VNGENENYIKYFRLYNOQIESTIPTIDD 937

QY 1189 TKVQEQVDPTPLAAATKFKVKDQIPVDGSGNIS-----DKWLASIP- 1232

Db 938 LEKIEFD-----LEITSKNIVEFRENILIEIKLSPKLYINEHYILEDLDKQISQIPH 993

QY 1233 -----LVIIHQMLRLSPVVKTIRELGLKTEQOQOQOQOQOQOQOQKAVKKEBEL- 1282

Db 994 NKKYEFDIKVPKNYLNKKNKVIK---VKLENHEPIRLDLKYINMAPIELQASDYBELL 1049

QY 1283 -EYINPKD-BFNILNPLTKAHLRLTSLNVLNNDPNYKIEDLVKIKNEAGDHQLAFSLRANN 1340

Db 1050 IELITKVDQBFKLN-LEYGYDKIENLFNID---LINNL-LRKSESVKNIQREFTIVSNS 1104

QY 1341 IKRLMNTPIFADYNPFPYFNEDWRSIDKYLNNKGNVSSHQOQAAAGNOGSGLTQRLMKN 1400

Db 1105 HKSLGSTSINLINL-----KNDFKEITEIPIGKQN-----QIQAN 1140

QY 1401 IKPETFPTALIALKORNTNLSNYSOKIMIKPKYLVERSIGVPMSTGLDGYTGSBQTK 1459

Db 1141 L-----LKEKSK---KSSSDILYILPVVFWVVIIG-----GLIWFKKKNQR 1180

RESULT 25

A42771
reticulocyte-binding protein 1 - Plasmodium vivax
C:Species: Plasmodium vivax
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 28-Apr-1995
C:Accession: A42771
R:Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
Cell 69, 1213-1226, 1992
A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
A:Reference number: A42771; MUID:92315338; PMID:1617731
A:Accession: A42771
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-2829 <GAL>
A:Experimental source: Belem strain, merozoites
A>Note: sequence extracted from NCBI backbone (NCBIN:108114, NCBI:P:108115)

Query Match 2.9%; Score 282.5; DB 2; Length 2829;
Best Local Similarity 19.4%; Pred. No. 0.00075;
Matches 423; Conservative 328; Mismatches 757; Indels 673; Gaps 111;

QY 49 LIDSVAPK-----SIANFTSDYQSVKALLNGKTFD-----PKSSEBTFDFVSKPD 94
DB 641 ILDKMAKVHYLKELLSLKSGSVYPTENNELLNTASYDNMEGFSAKKEKADNDINALYN 700
QY 95 -----FLTNNGRTVLEIPKKYQVVISSEPSDEDDKERR-RLGFLHLEKLE 137
DB 701 SVYREDINALIEVEKFTVENKESTLEMLKDBEM---EEKQDAKETPAKLVFVSDDKLT 757
QY 138 DGNIAQSAKFTYLLPLDMPKALGOYSIVDKNFNNLIH-PLSNFS-AQSIKPLALTR 195
DB 758 DVTYKMSAE-----VTNAGIKKELIAQKQFEN--VHKMKFEFSADFSTKFEALQN 805
QY 196 S-----SDFTAKLQFNQDELWYLEKFPDLEALKANI-RLQTAQ-----PSF 238
DB 806 SMQOYNOEGDAIEKHQNRSEK-----BEYFNKESVEEDLSGREETEEOEYTKHKNFSR 860
QY 239 EKGNIYDPPV-----YSFIERNPONQKEMASDLAQDOKTVRLYL 277
DB 861 RKGEISAEITNMREVINKIESQNYGVYIEKYPFLIGDQNEVSTAKALK--EKIVSDSLR 918
QY 278 TESPQAKTILDKYKDETFSSI-----DLKASNGTSLFANENDLKQDLVDLDD 329
DB 919 DKID-QYETFEKTEKSAVENTVSTIQSLKSAIDSLKELNGSI---NNCKYNTDIDL- 972
QY 330 VSDYFGQSETITSNSQV-KVPVASE-----RSLKDRV-KPKKQOQKPRIKFP 375
DB 973 -----RSKIKTLREEVQKMPKRGKCGENTTALLKSLRDKMGKINEKLNKDLGRLNSL 1025
QY 376 SLYEYDALSPYQELVSKPNSIKDLVNTATLARNLRFPSLGKYNFLPDDLASHLD---Y 431
DB 1026 DTKKEDLLKFPYS---ESKSIHLKSD-----QKGPDDPLNRID-EWEIDIKRDVDELNVN 1075
QY 432 YFLVSKAKI---KOSSTTKKLPTELPIKISLSSILGDOBPNIKTLFEKEVTFKLNDFRD 488
DB 1076 YQVISENKVTLFKNVSVT---YIEA-----MHSINTVAHGITSKNELKS 1119
QY 489 V-BIEKAFGLL-----YPGV-NBELEQAKAQAASFEKEKS---KGLKFPQOK----- 533
DB 1120 VKEVEDKLVLEQVEDYKKNPENENKQLEAIRGSMKLEKVINKHVSEMTQLESTANTL 1179
QY 534 EENSKAINQOEGLEEDB-----NITERLPENSPYOQOEN----- 568
DB 1180 KSNAGKGENHDELELNKTKGMRDIYEKJJAELKEGTVNLKADANEKANKVPEPE 1239
QY 569 -----AGLGASPDKPYMIKDQVQRYYLAKSQOELIKAKD-----YTKLAKLL 612
DB 1240 RNIIGHVLERITVEKDAGKVVEEMS---LTKIEKLIQETSDDSQNELVTSITKHL 1295
QY 613 SNRHTY-NISLRLEQQLFVNPRIPSRDIEKAFVLDKTEKNYQWQIYSSASFPVQNKW 671
DB 1296 ENAKGYEDVIKNEEDSIQLREKAKSLTETDENKKLVQVNMNLQSAIQGNAG-ISKELN 1354
QY 672 SLFGYRYRLGLD-----PKQTIHELKVLQKAGLOF---EG-----Y 706
DB 1355 ELKGVIELLSTNYSSILEYVYKKNSESVRFSQLANGFEFTKAEGEKNASARLAEBKLK 1414
QY 707 ENLPSPDNLEDL---KNIR-IKTPFQKDN-----FK-LSLLDFNNYDGEIK 750
DB 1415 EQIVKLDYDSDIDDKVKKIEGIEKREILUKMESALTFWEESEKFKQMCSSHWENAKGKK 1474
QY 751 APEFGLFLFLPKELRRNSNSGGQSNSPWEO-----E 784
DB 1475 I-----BYLKNNGDGKGANITDSQWEEVGNVSKAEHAFHTVEAQVQDKTAFCE 1523
QY 785 IISQF--KDQNLNQDLAQFSTKIWEKIIDGENE-----PDQNNRLQYKLLKDLQEB 834

DB 1524 SIVAYVTQDNLNFSLEMKVEKVKCEKK--NDEAKYSAKLKYD--GRIKARVSEN--E 1577
QY 835 SWINKTRNLNVTWVIGDKLKVKKPK-----NMLEAK-FQIINLOELLTAPTYS-AALSNN 887
DB 1578 RKISE-----LKEKAKYKEKESQINDVSTKSLQIDNCRCQQLDSVLNIGRVKON 1628
QY 888 WNYTODSGAKSTIIPFEEIAELDPKV---KEHGVADVQYQKPHYAIGPDNAGKPNQEVIR 944
DB 1629 ALQVFDSDAKMKSVLPISLGAESKSLDKVKAESYK-----NLETVQ 1673
QY 945 SSSRTIYLTSGSKLEADTIDQLNQAVK-----NAPLGLGSFYLDTERF 989
DB 1674 NEMSRINVEGSLTDIDKKITDIENDLLKMKQYBEGLLQIKENADRKSNFELVGSBI 1733
QY 990 GVFOKLATSLAVOHKQKEKTLPKLANNQGYTL--IHDKLLKPPVPIQISSPEKDFEGKL 1047
DB 1734 NALLDPSTSIIFKLKLEKDYDMTGDKNYGVKQNEIHGFTK----- 1774
QY 1048 NQNGSQNVNVTFGSIIIESPVFTNFOBDADLDQDQDQDSRQGNNSLDNQEAGLLKQKL 1107
DB 1775 -----SYNLIETHLSNATD---YSVTFEKAQSLRE----- 1801
QY 1108 AILLGNQFIQYQQNDKEIFEIINVEKVSLSFRVFP--KLA----- 1148
DB 1802 ---LAEKEBEHLRRREBEAIFLLNDIKKVSLLKLEMMKVKVSAEYEGMKRDHTSVSOLV 1858
QY 1149 ---KTLBNGKTIRVLSDETWSLIVNTTIEKTPENSAYPEVDTKWVE-----QYDPR 1198
DB 1859 QDMKTIIVDELKTLNDISE--CSSVLNNV-----SIVKVKESKHADYRRDANSYBEM 1910
QY 1199 TPLA-----AKTFVLKFKDOIIPVDGSGNISDKWLASIPLVIHQOMLRLSPVVKVTIRE 1251
DB 1911 VTLANYFLSDEAKISGHEFNAMKSNFKTDLE-----LEIFSVISNSNE 1955
QY 1252 LGLKTEQOQOQ---OQOQOQOQPKKA-----VRKE--EBELETVPNKODEF-----NI 1293
DB 1956 LLKKIEQSDNDVIQKERESEQLAKDADIYVNIKLNKNEFNKLEAKNEKEEVVSEKREA 2015
QY 1294 LNPITKA-----HRLTSLNVLNNDPNYKIEDLKI-----KNKAGDHLQALF 1334
DB 2016 LKRLSQVEGIRCFHFENFHL-----DNTBELENLKKOMVTIYRDKKSERSEGLQEMEN 2068
QY 1335 SLR--ANNIKRLMNTPIPTFA---DYNPPFYNYEDWRSIDKYLNNKNGVSSHQOQOAGGN 1388
DB 2069 EMWTTYSNTQLEGIVVSAGESKEIDIEKLENEEMRNISE-----KLTIDSQVIEWN 2122
QY 1389 QSGSLIQRINKNIKPEFTPTPALIALKDRNNTNLNYSOKIMI-KPKYLVYERSIGV--PW 1445
DB 2123 STIDELYLKLGKNCQAHWI--SLISVTANMKT-----SKKLIMINKENTKCKVDYIKON 2175
QY 1446 STGLDGYTSGEOTKGTSSSQOQGFDDQDFIQAALGLKNTY--HGKLGLS-----IR 1495
DB 2176 SSSTDGYV---ETLKGFGYSGKLTFFSSASEIVQADTYSVNFPAKHEKESLNAIRDIKKELY 2232
QY 1496 IFDPGNGELAKIKDASNKGEELKLSYD-----LFKNYLNKEYKK-----SPKIA 1540
DB 2233 LFHONSDDISIVEG-----GVQNMALYDKLNEKREMDLYRN-ISETKLKQHEHSTDFV 2286
QY 1541 KGMWNTIHPDQKQVYPNPQKLPENYLNLVQPKVTLNYSDDPITNLFVEPEGSDRGSGT 1600
DB 2287 KPMIELHKGWNETNNKSLLEKEKLSVNDHMSM---EAEIMKN-----GL 2330
QY 1601 KUKQVLOKQVNNYA-----DWGSAYLTFWYDK---NLIITNPQNVITANIA 1643
DB 2331 KYTPESQVQNNIYSVIEAEVKTLEBIDRDYGDNYQIVEEHKKQFSILIDRTNALMDDI- 2389
QY 1644 DVPIKDVKELEDNTKLIAPNITQWPNISGSKEKFKYKPTVFFGNWENENSSMNSQAQTPT 1703
DB 2390 EIPFK-----ENNYNLMEVN-TETIHRVNDYIEKIITNKLV-----QAKT-E 2428
QY 1704 WEKIREGFALQALKSSFDQKTRFTFVLTNAPLPLWKYPLGPFONGPNFKTQDRLVFPND 1763

Db 1296 LKQRIQVDDKRLKKNERNLHFTSNTERKRAVLEDOISYFVKORKQATDAIIASHKE 1355
QY 1525 FKNYLNEYKSPKIAKAGWTNHPDOKEYPNPNQKLPENYLVNQPWKVTLVNSDPFI 1584
Db 1356 VKKGELOQLAVELETRKTLNNDPAKSRQREBFENQRLKLELOKTLQOTQTSNPFK 1415
QY 1585 TNLVFEPEGSDRGSGTKLVQVTKQVNNYADWGSAYLTFWYDKNIITNQPNVITANIAD 1644
Db 1416 TRAIQBIENSYKRGMBELN--FOKKE-----FDKN-----KSLRYE 1449
QY 1645 VPIK---DVKELEDNTKLIAPNITOWPNI---SGSKERFYKPTVFFGNWENENSSMNSQ 1698
Db 1450 YFRKORDETERKESQVKVLKE--TQKANLLEAQANKLNIERTIDFKEKE----- 1499
QY 1699 AQTPTWEKIREGFALQALKSSFDQ 1722
Db 1500 -----LKAFKDKVDQ 1509

RESULT 27
D71623
erythrocyte membrane protein pFEMP3 PFB0095c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C:Accession: D71623
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Perteira, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: D71623
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2441 <GAR>
A:Cross-references: UNIPROT:O96124; GB:AE001371; GB:AE001362; NID:G3845092; PIDN:AAC7180
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0095c

Query Match 2.9%; Score 281; DB 2; Length 2441;
Best Local Similarity 20.6%; Pred. No. 0.00071;
Matches 349; Conservative 234; Mismatches 614; Indels 498; Gaps 85;

QY 74 LNKG-----TFDP-----KSSEFTDFVSKFDPLTNNGRTVLEIPKQYQVISEFS 118
Db 481 VNGKLLKSTPNPKRRNKLKERKQBLHKKNY-----KTYQKLE----- 523
QY 119 PEDDKERFRLGFLHKEKLE--DGNIAQSATKFYLLPLDMPKAAALQGYSVIVDKNFNNLI 176
Db 524 -----REKRENPGE-----PLNTPFIHVIRPSDLMKNGENKSA 557
QY 177 IHPLSNPSAQSIK-----PLALTR--SSDFIAKLQNFNQ--DELVVYL 216
Db 558 GHFPKPYQPTKGLYEBESHVAKDYOLEHBPPTKLPEYKGVSRVQLDNEVRDELPEYE 617
QY 217 EKFPDLLE-ALKANIRIQTADFSEKGNLVDPFVYSPINPQOKEWASDLNQDKTVRLY 275
Db 618 KGHVSREYQLDNEVRDELPE--YEKGHV-----SREYQLD-NEGFPSTLKEY 660
QY 276 LRTPSPQAQTIILKDYKYKQDETFLSSIDLKASNGTSLPANENDL--KQQLDVLDDVSDY 333
Db 661 DQTELAKGKDITNKPHEVDEYDQSEL-----AKGKDITNKPHEVDEYDQTEL 709
QY 334 FGQSETITNSQVQKVPASERSIKORVFKPKDQOKPRIEKFSLEYDALSFYSQLOELV 393
Db 710 AKGKVTNKEHENLE-----EYNETDLAKGKVTNKEPH---ESVDEYDQ--SELAKGKDIT 760
QY 394 SKPN-SIKOLVNTATLARNLF-----SLGYNFLPDDLA-----SHLDYFLVS 436
Db 761 NKPHESVDEYDQTELAKGKVTNKPHENLEYN--ETDLAKGKVTNKAHENLEYNED 818
QY 437 KAKIKQSSITKULFIEL----PIKISLKSSILGDQEPNIKTLFE-----KEVTFK---- 482

Db 819 LAKGKE--VTNKAHENLEYNETDLAKGKVTNKAHENLEYNEDTLAKGKVTNKAREN 876
QY 483 LDNFRDVEIEKAFGLLYPGVNEELEQAKAQRASPEK--EKSXKGLKEFSQO-----KEE 535
Db 877 LEEYNEDTLAKGKVTNK--ARENLEYNEDTLAKGKVTNKAHENLEYNEDTLAKGKEV 935
QY 536 NSKAINNOEGLEEDD-----NITERLPENSPIYOQOENAGLGPDPKPMIKDVQNO-RY 589
Db 936 TNKAHENLEYNEDTLAKGKVTNKAHENLB--EYNEDTLAKG-----KEVTNKARE 985
QY 590 YLAKSOIQELIKAKDYTKLAKLLSNRHTYNf-----ISLRLKEQLFDVNPIRPSRD 640
Db 986 NLEYNEDTLAKGKVTNKAR--ENLEYNEDTLAKGKVTNKAHENLEYN-----ETD 1038
QY 641 TEKAFVLDKTEKN--KYWQIYSSASPVFQNKWSLFGYRYLLGLDPQKQTHIELVKLGOK 698
Db 1039 LAKGKVTNKARENLEEYEEKDYMKNNELQNGS-----DGLKENAELKNK 1084
QY 699 AGLOPEGYENLPSPDNLEDLKNIRIKTFLPSQKNFKLSLLDFNN-----YVDGEIKAP 752
Db 1085 E-LRNKSGDGLKENAELKN--KELRNKGS--DGLKENAELKNKELQNGSGELKENAELKNK 1141
QY 753 EF-----GL-----PLFLPKELRRNSSGSGSONSPWEQEIISQFQDNLSNQDLAOF 803
Db 1142 ELQNGSGELKENAELKNKEL--RNKSGSELKEN-----AELKNKELQNGS----- 1186
QY 804 STKIWEKIIGDNEFDQNNRLOYKLLKDLQESWINKTRDNLTYVLDGKLVKPKNNLEA 863
Db 1187 -----EGL--KENAELKNKELQNGSGELKENAELKNKE-----LQNGSGELKENAEL 1233
QY 864 KFRQISN--LQELLTAFTYSAALSNNMNYDQSGAK-----STIIPEETAEALDPKVKERGV 917
Db 1234 KNKELRNKSGELKENVYNNNDLKNQNDLQNKDLSNKMKNKELLNKDLSNKMKNKELLN 1293
QY 918 ADVYQLKFHYATGFDNAGKFNQEVIRSSRTIYLTSGKSKLEADTIDQLNQAQVKNAPL 977
Db 1294 KDL-----SNEDMKNELLNKDIRNKDLKSIG-----NMEQNTGLKNTP- 1333
QY 978 GLQSFYLDTERFVGFKLATSLAQHKEKTLPKKLNDGYTLIHDKLLK--PVIPQISS 1036
Db 1334 -----SKGQNTGLKNTPNERQNTG-----LKNTPSEGOQNT 1366
QY 1037 SPEKDNFEGKLN-----QNGSQNVNVTSGSIIESPYFSTNFBQDADLDQ--DGQDD 1087
Db 1367 GLKNTPSEGOQNTGLKNTPNERQNTGLKNTPS-----EGQNTGLKNTPPIEQQN 1417
QY 1088 SRQGNNSLDNQBAGLLKQKLAILLGNQFTQYQOQNDKEIEFBIINVEKVSLSFRVEFKL 1147
Db 1418 TGLKNTPSEGOQNTGLKN-----TPNERQNTGLKNAANKGQNTGLKNTP 1463
QY 1148 AKTLBNGKTIIRVLSDETHSLIIVNTTIETKTPMSAVPEVDTKWVEQYDPRTPLAATKF 1207
Db 1464 SKGQNTG-----LKNTPNERQNTGLKNTP-----NERQNTGL----- 1498
QY 1208 VLKFKDQIPVDGSGNISDKWLASIPLVIHQOMLRLSPVVKTIRELGLKTEQOQOQOQOQO 1264
Db 1499 -----KNTPSEGOQNTGLKNTPNERQNTGLKNTPSKGQNTGLKNAPNERQNTGL 1550
QY 1265 -----QOQOQOQOQKAVRKEELET--YNPKDE-----FNILNPLTKAHLRILNLSNVNDPN 1314
Db 1551 KNTPSEGOQNTGLKNTPSEGOQNTGLKNTPNERQNTGLKNTPSKGQNTGLKNTPSKGQNTGL 1607
QY 1315 YKIEDLVKNINAGDHQALAFSLRANNIKELMNTPTIPADYNPFYFYNEDWRSIDKLYNNK 1374
Db 1608 ERQNTG-LKNTPSEGOQNTGLKNTPSKGQNTGL-----KN-----TPS-----EGQNTGLKNTPS- 1654
QY 1375 GNVSSHQQAAG-----GNOQSGLIORLNKNIKPETFTFPALIALKDRNNTLSNYSOK 1427
Db 1655 KNAPNERQNTGLKNTPSEGOQNTGL-----KN-----TPS-----EGQNTGLKNTPS- 1698
QY 1428 IIMKPKYLVERSIGVWPSTGLDGYIGSBQTKDQ-----TSSSSQOKGFDQDFIQAQLGKNT 1484

Db 1699 -----GGQONTGLKNTPNRQONTGLKNTPSEGOQN-----TGLKNT 1735
Qy 1485 --EYHGKLGSLIRIFDPGNEELAKIDASNKKGEKLLKSYDLFNKYLNEYE-----KKSP 1537
Db 1736 PNERQONTGL-----KNAANKGQONTGL-----KNTPNRQONTGLKNT 1775
Qy 1538 KIAKGWNIHPDQKEYPNPQKLPENYLNVLNQPMKVTLYNSDDFITNLFVPEGSDRG 1597
Db 1776 SEGOQONTGL-----KNTPSEGOQ-----NTGLKNT-----PSEGOQONTGLKNTPNRQ 1818
Qy 1598 SGTKLKQVIQKQVNN 1612
Db 1819 QNTGLKNAANKGOQN 1833
RESULT 28
S73852
hypotheetical protein MG218 homolog F10_orf1818 - Mycoplasma pneumoniae (strain ATCC 29342)
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: S73852
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105885; PMID:8948633
A:Accession: S73852
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1818 <H1M>
A:Cross-references: UNIPROT:P75471; EMBL:AE000051; GB:U00089; NID:G1674211; PIDN:AAB9617
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: Mycoplasma genitalium hypothetical protein MG218
Query Match 2.9%; Score 280.5; DB 1; Length 1818;
Best Local Similarity 19.6%; Pred. No. 0.00049;
Matches 310; Conservative .261; Mismatches 628; Indels 383; Gaps 66;
Qy 56 KPSTANFTSD-----YQSVKALLNG-KTFDPKSSSEFTDFV--SKEDFTLNNGR 102
Db 323 KASLANLTKEKERLSAEKDSFERUNRTALDNINRMOENALFAKHLEQQOYEFERKQES 382
Qy 103 VLEIP---KTYQVWISPEPDDKERFRLGFHLKELE-----DGNIAQSATKFIYLLPL 154
Db 383 LKLETEHKQIQKRIGEFKISEAKSEALLIQERELLEKRREIDDLTQASLEYEQRR 442
Qy 155 DMPKAALQGYIYVDKNPNLIIHPLSNFSAQSIKPLALTRSSDFIAKLQFNQDELWV 214
Db 443 NQ---VLKEKRVQOQHQNLL-VHAKKK-----LDQRHYLAEQKEI---DEEQI 485
Qy 215 Y-----LEKFF-----DLEAKNIRLQTADFS---PEKGNLVDPF 247
Db 486 FLKEKIAITERRELEKLYLVKKQDQKQENLDLIFEKQLROYQADFENEIEKQ--ELF 543
Qy 248 V-----YSFIRNPQ---NQK-----EWA-----SDLN-----QDQKT 271
Db 544 ASQSKLSQSFQTKNKEAELKQAKIAEDNAHLKQNKHHHADLEIFLEGFNHLQKEKH 603
Qy 272 VRLYLRTSPQAKTILKDYKDBTFL-----SSIDLKASNGTSLFANENDLKQDLVDL 327
Db 604 KLEARTQFDNRVLSLARSFKQKQAEKLVKQSLQSLTAAPNFKSQEBAVERDWMKRL--ANL 662
Qy 328 LDVSDYFGQSETITSNQVPVPAERSLDRVKFKDQKQKPRIKPSLVEYDALSPYS 387
Db 663 EKQKMLGDVHQVDPENSLNISKLAEREL--ATKFEKELEAAQKQSLDNNNNAGLKL 720
Qy 388 QLOELVSPNKSIKOLVATLARNLRFSLGKYNFLPDDLIASHLDVYFLVSKAKIKOSSITK 447
Db 721 QLDKLSLKTRELRELEASKERILDF-----YDESSRIADYESDLARLAEVKTLE 772
Qy 448 KLFIPLPKISLKSILGQDQBNINIKLFEKVTFKLDNFRD-----VEIEKAFGLLYPGV 502

Db 773 K-----NQQETAAS--BRELKVALEKLNQAKFLQIRKQQLLEIASV 814
Qy 503 NEELEQAR---KAORASFEK--EKSKKGLKEFSQOKEENSKAINN---QEGLEEDNIT 553
Db 815 KQLAQKANLLKQQAELDKQTELEAAFLQDQDKKELEKALHSVKSQBELLEERSFL 874
Qy 554 ERLPENSPIQYQOENAGLGASPKPYMIKDVQORY--YLAQSQIQELIK-----601
Db 875 -----LQKQREFAEHVAGFKQVHFKTQMQLSEFNKQOQSEQIKRETELKIAFAD 926
Qy 602 -ADYTKLAKLLSNRHTYINLSLKE-----QLFDVNPRIPSSRDEKAEVL 648
Db 927 LKKDY-QLFELOKNOEFQIEQKHKELELLAOKAELEKQELQKATALASQODQTVQAKL 985
Qy 649 DKTEKNKYWOIYSSASPVFONKWSL-FGYRYLLGLDPKQTIHELKVLGOKAGL-----701
Db 986 DLARQOHELELRQVAF-----NQASLSLNKQREQLTNQVVLHGLKXKHEKLTLDKRLLA 1041
Qy 702 -----QFEGYENLPSPFN-----LEDLKNIRIKTPLFSQKDNFKLSLLD 740
Db 1042 EKEKDQHKDAEINQREPKQFENEYADFQAKRELOELNQIR-----RNLEQSNASILK 1095
Qy 741 PNNYDGEIKAPFEGGLPLFLPKELRNSSNGSGSONSNPWEQEIISQFKD-QNLSNQ--797
Db 1096 KRN-----QTLDFALLRKVQHNTQTNRVQLNTQI---KEFLLEKGNFQKASDEAA 1143
Qy 798 -----DQLAQFSTKIW---EKIIGDENEFQDNRLQYKLLQDLQESWINKTRDNLWYT 847
Db 1144 LQKALLIKRLRSFASKLOLQREALIQKLEFDRD-----EQKSEINNNAKLQ--1192
Qy 848 YLGDKLKVKPNKLEAKFRQISNLQELLTAPYTSAAALSNMNNYQDSGAKSTIIFEEIAE 907
Db 1193 ---EQFKLEKQNFDEAKOKILIEFKQDQRLDVEKRLKQKLQKLSKLSYLYKNRAD 1249
Qy 908 LDPKVKKEGVADYVQLKHFHYAIGFDDNAGKFNQEVIRSSRTIYLTSGSKLEADTIDQ 967
Db 1250 LSQOOLQHKYANLLELK-----EKLOTAKRALDKKHAIYGVKWAQFVSE 1293
Qy 968 LNOAVKNAPLGLQSFYLDTER-FGVFQKATSLAVQHKQKEKTLPKKLNDGYTLIHDKL 1026
Db 1294 LRQEKQQL-LSAQKQVDKSLLEQNGHQHQLNLSSETKKKQSLHEDINK-----FDOR 1346
Qy 1027 KKPVIPOISSPEKQWFEGLNQ-NGOSQNV--NVSTFGSIIESPYSTNPFQEBADLQDQ 1083
Db 1347 RKEAVSILNSHK-----KLKQKEGELQGLQKLSLKTQIEQE-FSKLYQOREKLDQ 1399
Qy 1084 QODDSROGNSLNDQEAAGLLKQKLAILLGNQFIQYQOQNDKEIEFEIINVEKVSSELSRV 1143
Db 1400 RTLLSKLHRELKQNEATAHKNREVLEIN-----YKKELQRLTTEKSEFDNNKRLFEY 1455
Qy 1144 EFKLAKTLENGKTIIRVLSDETSLIVNTTIEKTPEMSAVPEVDTKWVEQYDPRTPALAA 1203
Db 1456 FRKIRNEIEKKEAHIKTVLEETQK-----KRLHVEAVKLHLQKQSIIS 1500
Qy 1204 KTFEVLKFKQOI PVDGSGNISDKWLASIPILVHQOMLRSLPVVKTIRELGLKTSQOQQOQ 1263
Db 1501 KGQELKEIKERVSRDISH--TNKQREBELSLLHQKLLQKLAEREIREINNKDSLLTQKI 1558
Qy 1264 QOQOQOQPOKKA-----VRKEEE-----LETYN---PKDEFNINLPLTKAHL 1303
Db 1559 QTAQKQKLEKEARILKLEKRAVEQYQAEITLKTNRADLEKNDKHLFPPLFKIN--1616
Qy 1304 TSLNVLNNDPNY-----KIEDLVKIKNEAGDQLAFSLRANNIKRLMNTPTIFFA 1352
Db 1617 -----GNDMNPYPYPWFPWQQOQED-----SSNQIRHLFEQQLQFM 1653
Qy 1353 DYNPFFTYNE--DWRSTDVKYLNK-KNVSSHQQAAGNQGSGLIQRLNKI-----KPE 1404
Db 1654 QQR---YENELTELRRQALLEKLDQIQLSLSAKKNDPEKVEQMQLKTEKQKLS 1710
Qy 1405 TFTPALIALKDRNNTNLSNYS 1426

Db 1711 AFDQKINALAEQINTOKAEHAD 1732

RESULT 29

D71870

hypothetical protein jhp0928 - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004

C:Accession: D71870

R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Werberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999

A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori

A;Reference number: A71800; MUID:99120557; PMID:9923682

A;Accession: D71870

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-2231 <ARN>

A;Cross-references: UNIPROT:Q9ZKK7; GB:AE001522; GB:AE001439; NID:g4155505; PIDN:AAD0650

A;Experimental source: strain J99

C:Genetics:

A;Gene: jhp0928

Query Match 2.9%; Score 279.5; DB 2; Length 2231;

Best Local Similarity 18.8%; Pred. No. 0.00072;

Matches 401; Conservative 324; Mismatches 736; Indels 673; Gaps 105;

Qy 53 VAFKPS-----IANFTSDYQSVKALLNGKTFDPKSSSEFTDFVSKFDPLTNNGRTVLEIPK 108

Db 429 IDFKPSSEVLKGAHKTRYKANIKAEILLKELQAKQE-----ILKGDYYAT-----L 474

Qy 109 KYQVVISSEFPEDDKERFLRGLHKLKLEDCNIAQSAKF--IYLLPLD---MPKAALGQ 163

Db 475 KEQEILAQFSGWGLESYFYKKAQHPPEFKELNALLTKDEFRAYLSARDAYVTPKLVIDS 534

Qy 164 -YSYVDKNNFNLIHPLSNFSAQSIKPLALTRSSDFIAKLQFNQO-----DELWY 215

Db 535 IYQGLDQLGFNN-----DNHPKEIFEPSLGT--GKPIAHAPSDKNRYFTIGELDPISAN 586

Qy 216 LEKEFDLEALKANIRLQTAQPSKEGNLVDPFVYVSTIRNPQ--NOKEWAS--DLNQOKTV 272

Db 587 LSKFLYPNQVITALENYQFYQYD-----AFVGNPPYGNHKIYISNDKELNESI 638

Qy 273 RLY-----LRTFSPQAKTILKQYKDKETFLSSIDLKASNGTSLFA 314

Db 639 HNYFLGKAIRKELKDDGICAFVSWFMDAKNPKWREHI AKNATFLGAILR----- 688

Qy 315 NENDLKQDLVDLLDVSDYFGGQSETITSNQVQKVPVPSERSLKDRVKFKKQOKPKRIEX 374

Db 689 -----PNSVFQATGAEVTS--DIVFFKKGVEKATNQ 718

Qy 375 PSLVEYDALSPYSQLELVSKPNSIKDLVNAATLARNLRPSLGNVFLFDD-----LAS 427

Db 719 FT-----KAMPYDKIL-----NSLDDDTLQALQNN-RFD-----SFIPSDQLKIVNAVAN 763

Qy 428 HLDY-----YFLVSKAKIKQSSITKKFLPIELPKISLKSSTILGQEPNIKTLFEKE- 478

Db 764 HFGFKQELQRWYEKIDTANFGYSTQDYKIIKDFIDKVGKNSINLNEQTLINEYFIHPEN 823

Qy 479 ----VTFKLDNFR-DVEIEKAFGLLYFGVNSELEQARKAQRASPEK-----EKSKKGLK 527

Db 824 ILGHLSELEKTRYRFTETNGEQIYKYDQLQALEDESLDLSQALKQAIKLPKDVYQVHKTTLK 883

Qy 528 -----EFQOQKENSKAINNQGLE--EDDNITERLPENSPYQOEAGLQASPK- 577

Db 884 TDVLIIDSSNERYQEVQKLIKNLRRRELKVDNLYFQLEQNNEM-----GIFLKPTKI 936

Qy 578 PYMIKQVQNRVYLAKSQIOELIKAK-----DYTKLAK-----LLSN 614

Db 937 NSKQVDSRLKAYFKIKMDALNDLTSAELNPLSDLELENKRAKLNLYVDVFKKFGYLNEN 996

Qy 615 RHTYNISLRL-----KEQLFDVNPRISSRDIE-----KAKFVLDKTEKNKYWQ 658

Db 1896 KQDLFKK-LNRGGVRVLIGSPAKMGVG-TNVQERLVAMHEDCPWRPDELLQMEGRGIRQ 1953
QY 1553 YPNPNOKLPENYLVNLQPNWKTLYNSDDITNLVFEPEGSDRGSGTKLQKQVQKQVNN 1612
Db 1954 GNIHLQNDPENF-----RMKIYRYA-----TEKYDSRMWQIETK-SK 1991
QY 1613 NYADWGSAY---LTFWDKNIITQNPVITANIA--DVFIKDVK---ELEDNTKLIAPNI 1664
Db 1992 GIEQFRNAHLGLNELEDFNMGSSNASEMKAEGTGNPLIIEVVKLRAEIKSEE----- 2044
QY 1665 TQWPNITSGSEKPKYKPTVPFGNWNENENSNMSQAQPTTWKIREGALQALKSSFDQKT 1724
Db 2045 -----SKYKAPNKEHYF-----NEESLKNASKL-----DYLKQELK-DLETLQ 2082
QY 1725 RTFVLTTNAPLWKYGLPGFQNGPNFKTQDWRLVQFQNDNDQIAALKRVQEDRPEKSSD 1784
Db 2083 RSVIIPHTHEIKLDL-----KNESKDYELIKVEPEPLKENASMSBELTHKKLKE 2134
QY 1785 KDKQKWKIKFVVIPEEMPNSGNIRFVGMQIQGNTLWLPVINSVYDF--YRG----- 1837
Db 2135 QNKQ-----IAEQ--NKEKLDIAKKQFASNLNTLP---VNEEDYKLEYKGFVVNA 2181
QY 1838 -----TGDSNDVANL--NVAPWQVKTIAFT 1860
Db 2182 YKTKYQVEFSLSPKIDPNIAYSLAIWFKITLST 2215

RESULT 30
C90538
hypotheical protein MYPV 2110 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: C90538
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: C90538
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3216 <R>
A:Cross-references: UNIPROT:Q98029; GB:AL445566; PID:g14089624; PIDN:CAC13384.1; GSPDB:G
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPV 2110
A:Genetic code: SGC3

Query Match 2.9%; Score 279.5; DB 2; Length 3216;
Best Local Similarity 17.6%; Pred. No. 0.0012;
Matches 412; Conservative 361; Mismatches 824; Indels 743; Gaps 107;

QY 5 KSTLLLAT--AAAIIGSTVFGT-----VVGLASKVKYRGVNPQTQGVISQLGLIDSVAF 55
Db 820 BETILVLTLEKSQLATEEYFSKSFDFEVPGWASSELNNKNDLEGVQ-----NTLVS 874
QY 56 KPSIANFTSDYQSVKALLNGKTFDPKSSSEFTDFVSKPDF-----LTNNGRTVLEIPKKY 110
Db 875 KRPIIPFVKDFD-----KSTLLVDVAKLEDVFKDYSVVHNGKI-----KF 915
QY 111 QVVISSESPEDDKRFRGLFHLKEKLEDCN--IAQATKFIYLLPLDMPKALQ----- 163
Db 916 EVISWSSNPLSE-EQIR-----NFEQSNTEITHSGTATV-----RISIGENTNKW 959
QY 164 --XSYIYD-----KNFNLLIHH-----PL--SNFSAQSIKPLAL----- 193
Db 960 SDTFTFVVEGFPSSRKADNQKIVNDYVDLPETDTTKAKPILLARNVNTLAPRSLTSHF 1019
QY 194 -----TRSSDFTAKLNQF--NNQDELWVYLEK----- 218
Db 1020 EINTNRNRRRVNVIKSVEDVKNDQTKAIVTLEASAGSGNERATKEYAYRFEGLNNLQ 1079
QY 219 ---PFDLEALKANIRLOQATSFEEKGNLVDPPVYSF--IRNPQNK-----EWASDL---NQ 267

Db 1080 EDIYNDIEATARRIN-----AAQRANNIELISSFRLADPRSKKLASEALASDLFFANS 1133
QY 268 DQKT-----VRLYLR-----TEFSGPOAKTIL----- 288
Db 1134 NSGTIFSANTDSSGAQDISKVRFTTIGTRDGSINSSTLVEATIFSPNQAQLINKAPEL 1193
QY 289 -----KDYKYKDETFUSSIDLKASNTSIFANENDLKQD 322
Db 1194 APFAGQVQKTNANAQTSDLRVIAFQONQEYEL--YKIVSLRAHNSTNRAIDGSEVSQ 1251
QY 323 LDVLLDLDVSY-----EGG--QSETITSNQVKE--VPASERS-----LKDVKFKKQDQ 368
Db 1252 ISVNIKQSGFTPKVYTRIFSGYLSNDKFAKOKVDYIISGRSRANLVIDNAIKGQKSVI 1311
QY 369 KPRIEKFSLYDALSFYSQLQELVSKPN-----SIMDLVNATLARNLRS----- 414
Db 1312 ELEPSDFS-PDFQAAGNLLEIGTITTKGLTNTAVVEVLVSAGNTTNSRYTKSYNVEIAG 1370
QY 415 -----LGKYNFLPDDLASHLDYVFL-----VSKAKIKQSSITKKLFIEL 453
Db 1371 FKEPGAENRIRKLKKY---INDLANRKNPVLLOSVDKENTNASDLKASDFEIVKNAGFSG 1427
QY 454 PIKISLKSSILGDOEPNI-----KTLFEKEVTFKLDNFRDVEIEKAFGLLYPG 501
Db 1428 DIQTIESIARKORNPEVALVTIKVVTGEGDKVAEDKYTIEIPGFASNERIRHILGLISAY 1487
QY 502 VNEELEQAKAQRASFEKSKKGLKEFSQKBEENSKAINNQEGLBEDDN----- 551
Db 1488 TSSMSEALYQRAKLTQNRKPNQVSLLAITQ---DQLSVNSNFTIDQLSNHSQLHFSVLS 1543
QY 552 ITERLPNSP-----IQOOENAGLQSPKPYMIKDVONQRY-----YLAQSQIQEL 599
Db 1544 VPKPBBETTAIITIRIQAPTSNAGVAD-----KDIATFVYSBESGYLTNGQVPNF 1595
QY 600 IKAKDYTKLAKLSNRHTYNIIS-----RLKEQLFDVNPRIPSSRDTEKAKFVLDKT 651
Db 1596 LKADQYAKKAENQRKPIMLTGDKNTKTPVRLTINSFTTQPTTEGLKLEITKVEPDPT 1655
QY 652 EKN--KYWQIYSSASPVFQNKWSLFGYRYLLGLDPK---QTIHELVLKQKAGLQFEG 705
Db 1656 NNNNAKVTITVAGSGINENRSYVIEHIERGFASEGKAVNEANVKEYIESDKA----- 1709
QY 706 YENLPSPDNLEDLKNIRIKTLPFSQKDNFKLLDFNNYDGEIKAPE--FGLPLFLPKKE 763
Db 1710 ---LPS--LVVGIEKERVEAQSISSSNLF-----IQPDPTKGLSLTIEKV 1750
QY 764 LRRNSS-----NSGGSONSNPWEQEIISQFKQNLNODOL----- 800
Db 1751 NVQOQSLIVEIRVSAGSGQNLASQTYNHTITGATQEQTKNYALLRKXVTLPGRRTPSLV 1810
QY 801 -----AQFSTKIWEKIIDENEFQDNRLQYKLLKDLQ--ESW-----INKT 840
Db 1811 EPOSRSVKTVSQTNDWKIDQPGSGEY--SSIQLSIKEIKELERDKSWAVVIVEATINGL 1869
QY 841 RDNLYWTYLG-----DKLV-----KP--KNLEAKFRQI----- 868
Db 1870 KDTYQYLLSGFATQAKKVIDLVQKWVKEQAQTKNPTLKNEVETQKANIIVNLNVNDF 1929
QY 869 ---SNLQELLTAFVTSAAALNNNNYVODSGAKSTIIPFEEIAELDPKVEKVGADVYOLKF 925
Db 1930 IIPTNLPSNISIDLLSVS-----EKSDSETD---QKILARLVKVSAGSGSDYE-EF 1977
QY 926 HYAIGFDDNAGKQNEVIRSSRTIYLKTSKSKLE--ADTIDQLNOAVKNAPLQSGOSFY 983
Db 1978 YVY---DYEG--FAERRVQGSIRAI---TNVIANVENRAQVFLKVTSSINKHPSEITKED 2029
QY 984 LDTERFGVFOKLATSLAVOHKQEKTLPLKLNNDGYTLIHDKLKPKVLPQISSSPKDWKF 1043
Db 2030 LDIPNSG-----VENIKVEITEVKIENDATTL--DLNRVTAIVRVSSIDQ----- 2073
QY 1044 EGKLNQNGSQSNVNV---STFGSIIESPSTFQEDADLDDQD-----QDDSRQGNNS 1094
Db 2074 --SITSHTATYESNIFGFTTRASYLIQOLINSNARPSVLQTKDGSQQVSVQSEERKKILPS 2131

Db 2516 RSHMDDLSENAKLDIELIQYREDLNQVITIKDSQQKQLLEVLQONKELENKY-AKLBE 2574
Qy 852 KLWKPKNN--LEAKF-::|||:::|||||:::|
Db 2575 KLBSEBANDLRRSFNALQEBEKDLSKETESLKVSISQLTQVTVALQEBGTLGLTHAQL 2634
Qy 874 -----LLTAFYTS-----AALSNNWNYQDSGAKTIIIFEIAELDPKVKEKGADV 920
Db 2635 KVKEEEVHRLSALFSQSKRIAELEEELVCQKEAAK-----KVGIEDKLKK----- 2682
Qy 921 YOLK-FHYALGFDDNACKFNQEVIRSSRTIYLKTSKSKLEADTTDLQNAOVNAPLGL 979
Db 2683 -ELKHLHDAGIMRNETETAEERVAELARDLV-----EMEQKLL-MVTKENKGTAQI 2733
Qy 980 QSF-----YLDTERFGVFQKLAATSLAVQHKKKEKTKPLKLNNDGYTLIHDKLK 1027
Db 2734 QSFRSMSSLQNGRDHANHEELKRKYDASLKAQLKEQGL---LNRE-----RDALL 2785
Qy 1028 KPVIPTISSPEKDWF-E-GKLNQNGSQNVNSTFGSIIESPY-----FS---TNFEQA 1078
Db 2786 SETAFSNVSTEENSLHLEKLNQQLLSKQDLHLSSLQLEDSTYNQVQSFSAKASLQNER 2845
Qy 1079 D-----LQDGODDS-----ROGNNSLDNQEAGLLKQKLAILLNQ 1114
Db 2846 DHLWNELEKPRKBEGKORSAAQPSTSPAEVQSILKKAMSSLQNRDRLLKE-----LKNL 2900
Qy 1115 FIQYYQN-----DKIEFEELINVE-----K 1135
Db 2901 QQOYLQINQBITELPHLKAQLQEQYDKTKAFQMBELROENLSWOHEHLQLRMEKSSWE 2960
Qy 1136 VSELSPFVEFKLAKTLFDNG-----KTRIVLSDETMSLIWNITTEKTPENSAYPE-- 1185
Db 2961 IHERRMKEQYLMALSDKQQLSHNLRLRSSSTQPLKVQYQRAQSPETSASPDSGS 3020
Qy 1186 ---VFDTKWVEQVDPRPLAAKTKVLUKF-K---DQIPVDGSGNISDKWLASIPLVTHQQM 1239
Db 3021 QNLVYETELL-----RTLQNDLSLKEHQELRLIQOLNSNPSQLLEKNLTSLQCDCSTS 3075
Qy 1240 LRLSPVVKTRELKTEQOO-----QQCQOQOQOOP-----QCKAV 1276
Db 3076 LR-----ENQOHYGDLLNHCAVLEKQVQELQAGPLNIDVAPQAPQKNGV 3120
Qy 1277 -RKEEELETYNPKDEFNILPLTKAHLRLTLNSLVNNDPNYKIBDLKVINEAGDHQJ--- 1332
Db 3121 HRKSDPPEELREPQOSFSEQ-----QQLCNT--ROEVNELRKLEERDQRVAEE 3168
Qy 1333 -AFSLRANNIKRL-----MNTPPI 1349
Db 3169 NALSVAEEQIRRLREHSEWDSSRTPI 3193

RESULT 34
F71607
DNA helicase II BRAHMA homolog PFB0730w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C:Accession: F71607
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Perce, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A>Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: F71607
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A;Residues: 1-1997 <GAR>
A;Cross-references: UNIPROT:O96239; GB:AEO01414; GB:AE001362; NID:g3845260; PIDN:AAC7193
A;Experimental source: clone 3D7
A;Genetics:
A;Gene: PFB0730w

Query Match 2.8%; Score 274.5; DB 2; Length 1997;

Db 1230 EFRGYKTHRLDGNIGLOE--RKIIIDQFNNNVEYKDEKQKQPCNCPGNNMNSGNENM 1287
Qy 1053 -----SONVNVSTFGSIESPYSTNFQEDADLDQDQDD--SQGNNSLD---NOBAGLKK 1104
Db 1288 NMSVNNMNSVNNMNSGNENMNSGNENMNSGNENMNSGNENMNSGNENMNSGNENMNSG 1347
Qy 1105 OKLAILLGNQTOYY--QONDKEIEPIINVEKUSLSFRV--EFLAKLTLEDG---KT 1157
Db 1348 NENMNSGNENKMTSSQNEKOTSQSV---KISELKKEBEINDFQIMDDKNVNGNQDAM 1404
Qy 1158 IRVLSDETMSLIWTTTIEKTPMSAVPEVDFTKWVEQYDPRTPPLAAK-----TKFVLK 1210
Db 1405 IFILSTRSGSLGNLTQAT--:::--:::--DFNPHQDIOAMCCHRIGKQNVVK 1454
Qy 1211 FKQIPVDGSGNISDKWLASIPUVIHOQMLRLSPVVKTTIREGLKTEQOQQOQQOQQO 1270
Db 1455 VPRFITLSGVEELVFK--KAQHKLSINDKVIQAGLFNK-----IYNDEDRQNKLDIIOR 1507
Qy 1271 PQKAV-----RKEELETYNPKDE-----PNILNPLTKAHLRLTSLN 1308
Db 1508 NQKNDMTHTPTNPLLLNYMKRNEELEYFLDPDKRYFGQYFSLLN-----TLNVEN 1561
Qy 1309 VNNDPNVKIEDLK-----VKNKAGDQO---LAFSLRANNIKRLMNTPTTFADYNP 1356
Db 1562 DSGQFTYMSDEKEENETVLSIIKKEKEEGEDDEENQORDENKEEDQED----- 1613
Qy 1357 PFYNNBDRSIDKYLANKGNVSHQOQAGNGSGLI---QRLNKNIKPETFTPALIALK 1414
Db 1614 ----KDDKDKOKDKKEBEEKKRXHILNNNNGIQGSSINEGVKEK-----ILDEY 1664
Qy 1415 DRNNTLSNYSKLIIMKPKY-----LYERSIGVWSTGLDYGISEQTKGTSSSSQ 1468
Db 1665 CNNTKCVKVSNERLIFPKRKHDTDDIQCDEKIKENEEDVDNIIQNKKN--RLRWECOK 1723
Qy 1469 KGFQDQFIQALGLKNTYEH---GKGLSIRIFDPCNELAK-----IKDASNKKGBEKL 1519
Db 1724 DKDDDI-----NSNIHWEKKIYMSEKDDTTKEYSDTHDPYINDKQVQDEEDYY 1776
Qy 1520 KSYDLFPKNYLNVEK---KSPKIAK-----GWTNIHPDQKEYPNPQKLPENYL 1566
Db 1777 -GFILKEENQDIEKILKSNKLINKDELPAFLFYDDTNDSPDK-----INLRSRKVIN 1830
Qy 1567 LVNLNQPKVT-----LYNSSDFTNLVPEPEGSDRGSGTKLQVIOQVNNYADWGA 1620
Db 1831 INLMQEEKLTQKQFLKIDSS--PNLLSSVE---KDLGRNKDIIVKSOMEHNN---DIT 1882
Qy 1621 YLTFWYDKNIITQPNVITANIADVFKDVKLEDNKTIAPI--TOWMP--NISGSKEK 1677
Db 1883 TLEVKDREBEIKHEHLETTNIISSNLND---LEINKTLTNEVHVSTKSPYNMRSSKR 1939

RESULT 35
E71619
RAD2 endonuclease PF0265C - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C;Accession: E71619
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Perte, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: E71619
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1516 <GAR>
A;Cross-references: UNIPROT:O96154; GB:AE001383; NID:G3845135; PIDN:AACT184
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PF0265C

Query Match 2.8%; Score 272; DB 2; Length 1516;
Best Local Similarity 18.6%; Pred. No. 0.00086;

Matches 317; Conservative 254; Mismatches 543; Indels 594; Gaps 82;
Qy 36 GVNPTQGVISGLI-----DSVAFKPSIANFTSDYQSVKKALLANGKTFDPKS----- 83
Db 2 GYKGLWSIVSPGVVRVNPFIPTGRIAIADVSIWLYELTYANNVKDLRN--KSPDNNSIFND 60
Qy 84 --SEFTDPSVKFDFLTNNGRV--LEIPKKY-----QVWISFSPDDDKERFRLQ 130
Db 61 LWIDSENISS--EIKTDNIKAHLFPFLRICKLYNIRPIFIDGNPPPELKRRTIQR 119
Qy 131 HLKLEKLEGNIAQSAKFIY-----:::--:::-- 150
Db 120 NIKKRYEKKFKTAELVYNYQRTLLNSMKSNNKNDNSNIEDKTNPNTKNTQNS 179
Qy 151 -----LLPLDMPKALGO--YSYI--YDKNFNLLIHLPSNFSQAQSIKPLATRSS 197
Db 180 NTQNSKNTPNKINADISKSSLIQIYDDIKEKDSLSNLSVEH--VGNVPV--SVKDV--LTICN 236
Qy 198 DFIAKLQ-----FNNQDELWVYLEKFPDLKALKANIRLOTADFSPEKGNLV 244
Db 237 DDLSKIKNKIFMTDFGPVLFLGEQDGMGTVENINKLD-----NRNKDNNLS 285
Qy 245 DPFVYSFIIRNPONQKEMASDLNQDQTVRLYLRTBFSPQAKTILKDYKYKDDTFLESSI-- 302
Db 286 YSINTYKVDVNNND--DDKDKKENINEVRRDQ-----KNYYVYKKNENINNIYL 334
Qy 303 -----DLKASGTSLPANENDLKQOL-----DVL 328
Db 335 DDDDEKEDIQNKGV---YNNDDIDEQIIRKHKMARKKYIESIPTKFGFLCMRPRVDII 391
Qy 329 DVSDFVGGQSETITNSOVKVPVPSERSLKORVFKKQDQOKPRIKBSFLYEYDALSIFYSQ 388
Db 392 DISNY-----NTEMLEI--SETLKVENKFK----- 415
Qy 389 LQEL--VSPKNSIKDLVNATLARNLRFSGYKYNFLFDDLAHLDDYFLVSKAKIKOSSITK 447
Db 416 -QHLNVLDENNSTPVVNNLLKNINY--KKN---DDLIEGGE-----K 452
Qy 448 KLFIELPIKISLSSILGDOEBENIKTLPEKVTYFKLDNFRDVEIEKAFGLLYPGVNEELE 507
Db 453 KSFILN---INVD-----CYSSNSRL 473
Qy 508 QARKQARAFE---KEKSKGLKEFSQOKBENSKAINNNOEGLEEDDNTERLPENSPIQ 563
Db 474 NDENIERGLINPITNDEKSIINNNNNNNNNNN--NNDNNNDNDVIIHNNNNMIY 532
Qy 564 YQENAGLAGSPDKPYMKIQVQNYLAKSOIQELIKAKDYTKLAKLSNRHTYINSLR 623
Db 533 -----DNKYNVB-----CSSHKNDSNGISKNINILEPNNLDTSIFLE 572
Qy 624 LKEQ--LPDYNP---RIPSSRDIEKAKFVLDKTEKNKYWQIYSSASPVFNKWSLFGYR 678
Db 573 GKDEYKVVYVNNKEEIRIPLFKEINKE--IFBKLPLKLYOILQD-----IKEEWYTNRIK 626
Qy 679 YLLGLDPKQTIHELKVLGQKAGLQPEGY--ENLPSDFNLEDLK-----NIR--IKTPLFSQK 731
Db 627 AKSKDDMDVFSQ-----VQLETYVYRMIKTDFEIEKLIKMAENIQSVGELLINK 677
Qy 732 DMPK-----LSLLDFN-----NYVDGEIKAFEGFLPL- 758
Db 678 DLSKNTDINIKDYNNVLQKKKKKKKKKFLNDILNTYNTFTESKYQDLYVKGESKEDIK 737
Qy 759 -----FLPKELRNS-----SNSGGSQNSNSPWEQEIISQPKDNLN 796
Db 738 NQIDFVTOCYRNDIIRTHDKSDIFKNIRKIDNNKYEIYNLELEQEEINEKKNYKN 797
Qy 797 QDQLAQFSTKIVEKIIDGENEF-----DQNNRLOYKLLKLOESWINKTRDNL----- 844
Db 798 NDSKNTFFLKI-----ENEFKKLLDDSDSIQFGSLLADIKE--YNYTADNLNNDN 848
Qy 845 -----YWT-----YLG-----DKLVKPKONLEAK-- 864
Db 849 KSLYEDGENFITRNEPITNEYEKNNIIVISDRQYNEEDIIPKDKIKKEKKNDDTSDD 908

Db 2706 MSLSQNSRDHANEELDELKRRKYDASLKEAQLKEQGL---LNRE-----RDALLSETAFS 2757
Qy 1034 ISSSPKDWFE-GKLNQNGSQNVNVTFGSIIESPY-----FS---TNFQSDAD----- 1079
Db 2758 MNSTENSLSHLEKLNQQLSKDEQLLHSSQLEDSSYNQVQSFKAMASLQNERDHLWNE 2817
Qy 1080 -----LDQGDQDSS-----RQGNLSLNQEQAGILLKQKLAILLGNQFOIYYQ 1120
Db 2818 LEXFRKSEEGKQSAAPSTPAEVQSLKAMSSLQNRDRLLKE-----LKNLOOQOYLQ 2872
Qy 1121 QN-----DKETEPEIINVE-----KVSELSF 1141
Db 2873 INQEI TELHPLKQALQEQYQDKTKAFQIMQEBELRQENLSHQHLDQLRMEKSSWEIHERRM 2932
Qy 1142 RVEFKLAKTLEDNG-----KTIRVLSDETMSLIVNTTIEKTPMSAVPE-----VFD 1188
Db 2933 KEQYLAISDKDQQLSHLQNLIRELSSSSQTPKLVQVQROQASPTSPDGSQNLVYE 2992
Qy 1189 TKWVEQYDPTPLAAKTKFVLKPK---DQIPVDGSGNISDKWLASTPLVHQMLRLSPV 1245
Db 2993 TELL-----RTQJNDSLKEHOKELRIQQLNSNFSQLLEKNTLSIQLCDTSQSLSR----- 3043
Qy 1246 VKTIRELGLKTEQOQ-----QOQOQOQOQOOP-----QKXAV-RKBE 1281
Db 3044 -----ENQHYGDLNHCVALEKQVQELQAGPLNIDVAPGAPQEKNGVHRKSDP 3092
Qy 1282 LETYNPRDEFNINPLTKAHLRLTSLNVLNNDPNYKIEDLVKIKNEAGDHLQ-----AFSLR 1337
Db 3093 BELREPOQSFEAQ-----QQLCNT--RQEVNELKLEERDQORVAENALSV 3140
Qy 1338 ANNIKRL-----MNTPI 1349
Db 3141 EQQIRRLHSEWDSRTPI 3159

RESULT 37
T30822
lmp1 protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30822
R:Jensen, L.T.; Ladefoged, S.; Birkelund, S.; Christiansen, G.
Infect. Immun. 63, 3336-3347, 1995
A:Title: Selection of Mycoplasma hominis PG21 deletion mutants by cultivation in the pre
A:Reference number: Z18884; MUID: 95369882; PMID: 7543881
A:Accession: T30822
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1365 <JEN>
A:Cross-references: UNIPROT:Q49525; EMBL:U21962; NID:G790243; PID:G790244; PIDN:AAA81013
C:Genetics:
A:Gene: lmp1
A:Genetic code: SGC3

Query Match 2.8%; Score 270.5; DB 2; Length 1365;
Best Local Similarity 20.1%; Pred. No. 0.00086;
Matches 328; Conservative 235; Mismatches 565; Indels 501; Gaps 80;

Qy 283 QAKTILKDYKYKDTFLSSIDLKASGTSIFANENDLKQDLVDLLDVSDFYFGQSETIT 342
Db 41 QKENLLQNFKKLQQLNELGYKLVINEINVFHEQVLOGSL-----KIN 84
Qy 343 SNSOVKPVASERSLKDVRVFKDQKQPKRIEKFSLYEYDALSFYSQQLVSKPNSIKOL 402
Db 85 NKSETKAIBETLKDAILTISKIKQINQKEL-EF-----AKFNEIKDK 130
Qy 403 VNATLARNLRFSLGKYNFLFDDLASHLDYFLVS--KAKIKQSSITKKLFIPIKLSL 460
Db 131 LQEVYKKNEL--SKQYEHYKQNIENELNKTPTISLESTLIEIQNATNNL-----IKLNE 183
Qy 461 SSILGDQEPNKTILFEKE-----VTFKLDNFRDVEIEKAFGLLYPGVNELEQ 508
Db 184 ST---KSKONIDNINAKELKASTSQANQLLPQLSDN--DSEITAKA-----KKSUDAEIKN 234

Qy 509 ARKA-----QRASFPEKESK-----KGAKESQOKE-----ENSKAINNOEGLEDDNI 552
Db 235 ANQAVASNTASMQSASKSLDAKVAIBITKKUETFNKDKAEKFNELKQTRNQ--IQEFINT 292
Qy 553 TERLPENSPIOYQOENAGIGASPDKPYMIKQVQORYYLAQSIOQ-----ELIKAKDYTKL 608
Db 293 NKQNPYSELISQ-----LTSKSDSKNSVTBSSN-----KSDIESANTELAQ-----L 336
Qy 609 AKLLSNR-HTVNISURLKEQLFDVNPRISS-----RDIEKAFVLDKTEKNKY 656
Db 337 AKANADKVQADNLAKSIKEQL-----NNSVSNANTLSAKLTDKONTIOQAKTELEKEVQKAD 393
Qy 657 WQIYSSASPVFQNKWSLFGYRYLLGLDPKQTIHELKVLGOKAGLOFEG-YENLPDSFN- 714
Db 394 QAIKSNNTASMQSASK-----SLDAK--VAIBITKKLETFNKDKAEKFNELKQTRNQ 442
Qy 715 LEDLKNIRIKTFLFSQ-----KDNFKLSLDFNNYDGEIKAPEFGLPLF----- 759
Db 443 IQEFINTKNPNYSELISQLTSKRDSKNSVTDSSNKSIDSANTELKQALAKANADKVQ 502
Qy 760 -----LPKELRRNSGGGSSNSPWEQELISOFKQONLSNQDQ-LAQFSTKIWEKIIGDE 815
Db 503 ADNLAKSIKEQLNNSVSNANTLS-----AKLTDKONTIOQAKTELEKEVQKAD 550
Qy 816 NEFDQNNRLOYKLLKDLQESWINKTRDNLVYTLGDKLVKPKNNLEAKFROI-SNLOELL 875
Db 551 QAIKSNNTASMQSASKSLDAKVAIBITKKLE-TFNKDK-----EAKFNELKQTRNQI 600
Qy 876 TAPYTSAAALSNNWYQ-----DSGAKSTIIFEEIAELDPKVK-----KV 916
Db 601 QEFINTN-KNPNYSELISQLTSKRDSKNSVTDSSNKS-----DIESANTELKQALAKA 653
Qy 917 GADVYQLKPHYAIGFDDNAGKFNQEVIR---SSRTIYIKTS-----GSKLEADYI 965
Db 654 NADKVQA-----DNLAKSIKEQLNNSVSNANTLSAKLTDKONTIOQAKTELEKE-I 703
Qy 966 DQNLQAVK-NAPLGQS--FYLDTERFGVFQKAT-----SLAVQHKQKETLPK 1012
Db 704 QKANOAKITNNTASMQSASKSLDAKVAIBITKKLETFNKDKAEKFNELKQTRNQIQEFINT 763
Qy 1013 KLNNDGYTLIHDKL-----KKFVIPQISS9-----P 1038
Db 764 NKNNPNYSELISQLTSKRDSKNSVTDSSNKSIDSANTELKQALNTAKAKSSIDNELRP 823
Qy 1039 EKDWFEGLKLNQGSQNVNVTFGSIIESPYSTNPOED---ADLDQGDQDSDRQGNNSL 1095
Db 824 LKNDLQSKIEEFGPIRNTNFSWISSKLETT--KNKLABELTKADAIKNPNSSSKQALKDS 881
Qy 1096 DNOEAGLLKOKLAILLGNQFOIYYQOQNDKEIFEIINVE-KVSELSFRVEFKLAKTLEDN 1154
Db 882 SQO-----VQK-----LGNELL-----KTIIEFGKVKETKNSNIGYRL-FKLAQAEQFN 924
Qy 1155 GKTIRVLSDETMSLIVNTTIEKTPMSAVPDPVDTKWVEQYDPTPLAAKTFVLKFKDQ 1214
Db 925 -----NSVDVK-----LKNWEEK---QTLSSKKQL----- 948
Qy 1215 IPVDSGNTSKWLASIPLVIHQOMLRLSPVVKVTIRELGLKTEBQQOQOQOQOQOPOKK 1274
Db 949 -----GNQSTK-----DYLTLQSLTSMST-----QESTIKK 973
Qy 1275 AVRKEBELETYNPKDEFNINPLTKAHLRLTSLNVLN-----DPNYKIEDLVKIKNEAGD 1329
Db 974 VI-----VNIQAHIRNNLSQYRLKADKLIANMKRGYGDVKVIGISLQKQDMLMD 1023
Qy 1330 HOLAF--SLRANNIKELMNTPTIFADY--NP-----PFYVNEWMRSIDKLYLNKGNVSSHQ 1382
Db 1024 SVLSVDDSLKDDFNKALR---VLVDGYTKNPPVSSWFINKNPSIENYQNLRLNLVREN 1080
Qy 1383 QAAAGGNGSGLIQRLNKNIKPFTFPALIALKDRNNNTSNYSDDKIMIKPKYLVERSIG 1442
Db 1081 EIL-LDKANDLDKRAETIK---FV-----DENINSLDQRAKRLKQELINAKN-- 1124

QY 1443 VPMSTGLDYGISGQTKDGTSSSQKQDQDFIOALGLKNTYHGKLGSLIRIFDPGNE 1502
Db 1125 -----DLNFTLNHQKQFTAKDITPK-----ISLENKLEINQYLLPIIK-----EKA 1169
QY 1503 LAKIXD-ASNKGBEKLKSYDLFPKNLYNEVEKSPKIAKGTWNTHPDQKEYPNPQKLP 1561
Db 1170 VSKISIEKNKKLEDIIRS-----NFYLWEKVEINKYISELNTNKOVELRS----- 1215
QY 1562 ENYLNLVLPQWVTLNNSDPITNLFVEPEGSDRGSGTKLKQVTKQVQNNYADWGSAY 1621
Db 1216 -----NINENKWE-----SIKOTLDNLNKENVS-----LLKEVI-----INNSNAQYSI-- 1256
QY 1622 LTFWDKNIITNQPNVITANIADVPKIDVKELEDNTKLIAPNTOMWPNISGSKEKFKYKP 1681
Db 1257 -----NRILSTVPE-----FIKVAQITRSN-----NLRSLAEIQPKP 1288
QY 1682 TVP-----FGWENENSSMNSQAQPTWEKIR-----EGPAL 1713
Db 1289 TLLDIDKNLKVKKTLDNKTLNSDNIAKNEKIRLLVDRSELKESLRWFDQNENISD 1348
QY 1714 QALKSSFDQ 1722
Db 1349 QRIKNKLD 1357

RESULT 38
F90608
ABC transporter permease protein [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: F90608
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galieson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: F90608
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2599 <KUR>
A:Cross-references: UNIPROT:Q98PE9; GB:AL445566; PID:g14090189; PIDN:CAC13947.1; GSPDB:C
A:Experimental source: strain UAB CTIP
A:Genetics:
A:Gene: MYPU_7740
A:Genetic code: SGC3

Query Match 2.8%; Score 270.5; DB 2; Length 2599;
Best Local Similarity 19.7%; Pred No. 0.0021;
Matches 384; Conservative 269; Mismatches 689; Indels 607; Gaps 95;

QY 39 PTQGVISQLG-----LIDSV--AFKP-----SIANFTSDYQSVKKALLNGKTFD 80
Db 699 PRSDVIKEINNQGTFELFDSIYBAFRKSDLVTRGFISKANLDMYGPFKALYNNKPHS 758
QY 81 PKSS-----EFTDFVSKFDELNTNGRTVLEIPKQYQVIVSEPSPEDDKERFLGF 130
Db 759 VLSGGSINSNIYFKMFFDWISEIN--KENGNTF-----FNDVILDFTKE----- 800
QY 131 HLKKELDGN-IAQSATKFIY--LLPLDMPKAALQYSVIVDKNFNNLIHPLSNFSAQS 187
Db 801 -LKQIKGGGATTOQKQDFIVREFKFKDKIFGSLSTTKLFSALPLNLEVVDIYTIYKDP 859
QY 188 IKPL-----ALTRSSDFIAKLN-----QFNQDELVYLEKFFDLEALKANIRLQT 233
Db 860 VKFTGPEVIVKIDFTKATSEIGLWYEQNKQKQDNKYILLSEIDVLLSI-----LDS 914
QY 234 AD---FSEKGNLVDPPVYFTRPNQKQWASDL-----NODOK----- 270
Db 915 ADQNKVRLGMKIVSEIDFNKFLNPEEQDSIYSRFFKTHGENQDVKNPFNKLDSKEGQE 974
QY 271 -----TVRLYLRTEFS--POAKTILDKYKDETFELSSIDLKAS 307
Db 975 KYSNVSKGLVELISFLDIKSFVARLKQNSEFKLLPKIPNNPKYKTNIS--ISSKNLWAS 1031

QY 308 NGTSLFANENLDKQLDVLDDVSDYFGGOSÉTTITSNQKVPVPASERSLKDRVKFKDQ 367
Db 1032 FLDSIFHGQSD-----ENGVLGINNDAKISILVFMFNISDAGEFLKIS 1076
QY 368 QKPRI-----EKFSIYEVDALS-----FYBQLQ-----ELYSKPNISKD--LVN 404
Db 1077 DKPLYIPSPKDDKKVGLGDISVLQSLPOLFAGQVDKSKVEFTQLEILNQLDOIKNLSLQN 1136
QY 405 ATLARNLRFSLGKY-----NFLFDD--LASHLDYYFLV-----SKAKIQSSITK 447
Db 1137 RNLSQSVELLKKVLPLEQDFNFEDKNVIEHIDANLYLTNQPFNSTNSLVNFSITAK 1196
QY 448 KLFIELEPIK---ISLKSIIIGDQEPNITLFE-----KEVTFKLDNFRDVEIKAFGLL 498
Db 1197 RL-LDAPSNHEAVNLASKQLKDSLGSVYEFSELSSGLKHPDYKLG-----KKAISLL 1248
QY 499 YPGVNELEQAKQAORASFEKESKGLKBSQOKEENSKA---INNQEGL--BEDONI- 552
Db 1249 ALFI-----RLRKESLSIENYQNDYKKIFDFINTKSDDENAFVNVVNSKTLTKVDONDFD 1303
QY 553 -----TERLPENSPIOYQOENAGLGLASPDKPYMIKDVQNRQRYLAKSIOBELI 600
Db 1304 IAGFISGKGLASTSRL--ANELFEKEAEN-----YKNKEIQFLDSLITNEARKLIE 1352
QY 601 KAKDY-TKLAKLNSRHTYINISRLKEQLFDVNPRIIPSSRDIEKAKFVLDKTEKNKYWQI 659
Db 1353 ENKEYFIQIAAILLSPKEYH-----PTLLNI-----FEYEDIHRNNPFK- 1391
QY 660 YSSASVPQNKWSLFGYYR-----YLLGL-----DPKQTIHELKVLQKQAGLQ 702
Db 1392 YSITNPVFENKIALNELNKQASAINOSKIYISGLSYALVDEPLNI-----LLPQVALW 1446
QY 703 FEGENLPSPNLEDLKN-----IRIKTLPQKONFKLSLDD-----FNNYVD 746
Db 1447 FTSLNTPENSONKDNKSNLAYLITNKVNLGRKNFQIQNFVSFVLGNSNEGVPNN-PE 1505
QY 747 GEIKAPFGLPLFLPKELRRNSNGSGNSNSPWEQEI-----ISQFKD----- 791
Db 1506 NDAKS-----NLFVDLDYLELIGINENSRKKID--EQELVFGINFKKTIYEFIDAVTYPH 1558
QY 792 --QNLNQDQLAQFSTKIWEKIIGDEN-----BFDONNRLOYK--LLKDLQESNI----- 837
Db 1559 TLDNVVFNVDVASYLVKVNQAYLVNKNKTIYTGPIPTNKLNFLELLKNKDESNLNVNG 1618
QY 838 -----NKTRDNLVYTLGDKLVKPKNNLEAK-----FROISN 870
Db 1619 SQFIIVGTDITTSVLY-----PVDENNQVNTSQQALVYVYNQKGFBRVKTFSESPN 1670
QY 871 LQEL-----LTAFTYSAALSNNWYVQDSGAKSTIIFEEIAELDPKVKKEVG-- 917
Db 1671 IKEYILVKABPKKERLLADQIESYSQOTD--QSVGIKKTLYLNNEIDPINPERSIRVATP 1728
QY 918 ---ADVYQLKEH-----YAGPDDNACKFNOEVISSRTI--YLTGSGKSKLEAD- 963
Db 1729 RLIISSVSNFNFYIIIIILITLVAISIIIFVHY-----IKNLSLVIGILISQYSPQLISL 1784
QY 964 --TIDQLNQAVQNAPLG-----LQSEYLD-----TERFVGFKLA----- 996
Db 1785 SMTVPAMFTAVFGVFGVYVGNVNLQGLMNVFSNFWTLERHTEKFSIVSLMVAVFLPFLG 1844
QY 997 -----TSLAVQHKKQEKTLPKKLNNDGYTLIHDKLKKPVIPOISSPEKDFEGKLNQ 1049
Db 1845 MSILIIILTSMLWRKKPIDLM-----SGITAI-----KVGS-----FSRINR 1882
QY 1050 NGQSONVNSTFGSIIIESPVSTNFQEDADLDQGDQDSRQGNNSLDNQEAGLLKQKLAI 1109
Db 1883 LFSKRNKTFESALLVNSFEK-----LLSLAFAI 1912
QY 1110 LLGNQFIQYYQONDKEIEFEIINVEKVSSELSFRVEFKLAKTLEDNGKTIIRVLSDETWSL- 1168
Db 1913 TITSTIIMFGISINKVFSKSIDSTYKNRHSFKL--NLATPTLEGGLTKLNYDEIKNSL 1970
QY 1169 -----IVNTTIKTEPMSAVEVEDTKWVEQYDPTPLAAKTKFVLKFKDQIIPV 1217


```
Db 1971 YVPLGTPGSAIVDGYFYFKPGWSNAING--DRSQKNGDPKDYL---PHVFAFEMNIIIV 2025
Qy 1218 DGS--GNISDKWLASIPVHIQOMLRSLSPVVTIARELGLKTEQOQOQOQOQOQOQOQPKKA 1275
Db 2026 EGAIREFNLWLIYNSLP---DTQKFR--VDKIVSEMGLLEOTQFDETSFY-----KL 2073
Qy 1276 VRKEELEY-----NPKDEFILNPLTKAHL-----TSLNVNDPNYKIEDLVKIK 1324
Db 2074 IQNPENGNYGVSVDPKNYFYFDLQGVRGFFYMRWNAIENTYNHEPITTSRDI--FAD 2132
Qy 1325 NEAGD--HQLASLRANNIKRLMNTPIITFADYNPF-----FYNEWRSDIKLYNNKG 1375
Db 2133 DYYGDPNKNISLR--NV-----YRQFLVRAITKLHSENR--IAKLEERG 2177
Qy 1376 NVSSHQQAGAGGQSGLIQRLNKNIKPETTPA-----LIAKL 1414
Db 2178 TQSPFREDYLSLGGVVFPQRPD-----ETYSVASANFEGTTRSLNIYGYKSDSKFIKII 2232
Qy 1415 DRNNTWL-----SNYSKIIIMPKY--LVERSIGVPMWSTGLDGYIGSEOTKDGTSSSSQ- 1467
Db 2233 DENDVLMQBIAKFETKTIDGKKVPVWVNTVSSKKYKLGIGSTFGVSLNKATRYSHQI 2292
Qy 1468 -QKGPQDPT-QALGLKNTFYHGKLGLSIRIFDPGNELAKIDASNKKGEKLLKSYDLF 1525
Db 2293 KOENFDNNVIFVVGINPTFINEFITSOEIVNEITDLNKI-----ALRENOEIP 2342
Qy 1526 KNYLNEYEKSPKIAKGTWNIHPDOKEYPNQKLPENYLVNLVLPQWKTLYNSSDFIT 1585
Db 2343 NGILTSRDKPAQLI--GSSSIYASGWP-----AITNPEFK--GNESAIYD 2385
Qy 1586 NLFVPEGSDRGSGTKL-KQVIOKQVNN 1613
Db 2386 AIPDESGLRSVSGNLSEQQILRWLSN 2414

RESULT 39
T28317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C:Accession: T28317
R:Afonso, C.L.; Tuleman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: Z20484; MUID:99102612; PMID:9847359
A:Accession: T28317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1127 <AFO>
A:Cross-references: UNIPROT:Q9YVT6; EMBL:AF063866; NID:94049647; PIDN:AAC97677.1; PID:94
C:Genetics:
A:Note: MSV156

Query Match 2.8%; Score 270; DB 2; Length 1127;
Best Local Similarity 19.8%; Pred. No. 0.00069;
Matches 250; Conservative 187; Mismatches 426; Indels 398; Gaps 62;

Qy 533 KEENSKA-----INNQELEDNDITERLPEN-----SPIQYQOENAGLAGSPD 576
Db 22 KYENKRVSLDIINSLEYILNN---IKFSDKITNEIKYNYKIVKIFYMHNQ----- 70
Qy 577 KPYMKDVQORVYLA--SQIOBLIKADYTKLAKLLSNRHTYNTSLRLKEOLFVNPR 634
Db 71 --PKINDYNIILQLYBYNNEINKIKEN-----KPKCKNPLYNTYKKKLYIYDLYE 122
Qy 635 IPSRDIIE---KAKFVLDKTE--KNKQWQIYSSASPVFQNKWSLFGYRYRLGLG---DPK 686
Db 123 EKKOKELVINTEQKNAVVDKINDIKNNVNNIHSNDNETIITGTETLIDILNKLKLVSSDEK 182
Qy 687 QTIHELKVLGQKAGLQEGYENLPSDFNLEDLKNIRIKTLPFSQKNFPLSLDLFNYYD 746
Db 183 QLIEQIYKNINNKIEIFNIDNVQKEIN---KKQDELNKLDESKKEF----- 227
```

```
Qy 747 GEIKAPFGLPLFLPKELRRSNSGSGSONSPWEQBIIISQFKDQNLIS-----NQDOLA 801
Db 228 --IKRQE-----ELNKTIIDK-----QBELIKKLNKKEINFINIDEKQKLLD 266
Qy 802 QFSTKI---WEKIIIGDENEFQNNRLQYKLLKDJQESWINKTRNLNLYWYTGDKUKV--- 855
Db 267 QINSKINTLNENINGVMNLYTETK---NKISNLQNEILNK--DSTIKS-LDEKQKLLDE 319
Qy 856 --KPKNNLEAKFR---QISNLQELLTAFYTSAAALSNNWYVQDSGAKSTIIFEIAEIDL 909
Db 320 LDKNNINNTSYLNKSNKTIQNIQQLLESLLD---FNNANI-----NINELK 363
Qy 910 PKVKEKVGADVQLKFHYAIFDQDNAGKFNQEVIRSSSR--TIYLTGSG---KSKLEAD--- 963
Db 364 SKIKL-----FDNDIQKLANDITEQNNKIDTFNNSTRIFKEKLDTEYK 407
Qy 964 TIDQLNQAVQNAPLQGLQSYLDTERPFGVQKLAISLAVOHKQKE--KTLPKKLNNDGYTL 1021
Db 408 KIDD-----IKNNNL-----QKLEESYKKIDQETYYKKNKINKEYNDIIEEL 448
Qy 1022 IHDKLUKPVIFQISSSPKDNFEGKLNQNGSQNVNVSTFGSIIIESPFSTNFQEDADLD 1081
Db 449 KNNNLQKLEENKIKIDQETYYKKNKINKE-----YNDIIEEL--NNNLQKLEEN 496
Qy 1082 ODGQDSDSGNNSLDNQEGALLKQKLAJLLGNQFIQYQONDKEI-----EPE----- 1130
Db 497 KINIDKLTKLNDIESNTE-----LFWKLNISDFKDKRSREIAKLNTEYEQLRKDL 547
Qy 1131 INVEKVSLSFRVEPKLA--KTLEDNGKTIIRVLSDETMSLIVNTTIEKTPEMSANVEPVD 1188
Db 548 ENINKTNELMKLSDNKLSSLEQLYDSKKNIL---DGIDKIYNSLKEKNDK---IDEYES 600
Qy 1189 TKWVOYDPTPLAAKTKFVLKFKDQIPVDGSGNISDKWLASIPLVHQQMLRLSPVVK 1248
Db 601 N--IEKFDIYNVI--ENKFI-----GNLD-----SLINKI 626
Qy 1249 IRELGLKTEQOQOQOQOQOQOQOQKAVRKKEELETYNPKDEFNLTPLTKAHLTSLNL 1308
Db 627 I-----NNDQFKEYINSKIDSKSNELSTMP--DDIFNAKQNTA-----SITNN 667
Qy 1309 VNNDPNYKIEDL--VIRKNEAGDHQALFSLRA-----NNIKRLMNTPI--TPAD----- 1353
Db 668 IENISN-KIKOLNEFIISNEDSSKELLDEIRKYKQOFDKIDAMNTEVKSFENTLQKID 726
Qy 1354 -----YNPFYFNEDWRSIDKYLNNKKNVSSHQOQOAGNOGSGLIQRLNKNIKPE 1404
Db 727 SIKSNINELTNAYDIINTKANLDLDDKLNLYDSSEFKNLY-----NNASDLDDTIQKN--- 777
Qy 1405 TTPPALIALKDRNNTNLSNYSDKIIMIKPKYLVERSIGVPMWSTGLDGYIGSEOTKDGTS 1464
Db 778 -----NDEKVKQLNLEYLK-----NKNQ 795
Qy 1465 SSQOQKGFDDFIQAL--GLKNTFYHGKLGSLIRIFDPGNELAKIDASNKKGEKLLKSYD 1523
Db 796 SIEINDIVNNFIKELIKFNNTETNKSINELLNTDNDINDIKIFKLYKELNKISTNNLLK--- 852
Qy 1524 LFNQVLYNEVEKSPKIAKGTWNIHPDOKEYPNP-----NOKLPENYLVNLVLPQWKT 1577
Db 853 IYKNEIDNVDNEK-----LSIVIENTLOFINSFUSIEFNOGSIITSHINFLN-----TL 899
Qy 1578 YNNSDFIT--NLFVPEGSDRGSGTKLKOVIQKQVNNNVDWGSAYLTFWYDKNIIITNQ 1635
Db 900 AGINDVNLKLNKIMADTTRGD--TNIRDEIKNQIISSE----- 936
Qy 1636 NVITANIADVFKDKVKEL--EDNTKLIAPNITOWMPN-----ISGSKEKFY 1679
Db 937 NIKSQKFNEKNEKDLKLI SFNDKLNKYNISAGYTEYNNIEHCLKLYLIAVSDQBYRY 996
Qy 1680 K 1680
Db 997 K 997
```

Query Match	2.7%; Score 267; DB 2; Length 1830;
Best Local Similarity	19.2%; Pred. No. 0.0018;
Matches 380; Conservative 274; Mismatches 739; Indels 582; Gaps 94	
Qy	30 SKVKYGVNP--TQGVISQLG-----LIDSVAFKPSIANFTSDYQSVKKALL- 74
Db	86 TKINYLAIGDSITAGFNSLGEWAPGRYDPTNKISGLSPSPFIAQYINKVFNRLASYE 145
Qy	75 -----NGKTFDPKSSSEFTFVSKFD-----FLTNGRT 102
Db	146 NLGITGSRIEDWYMLGAGSDEYFQNGEYRFFDSKQMDQOKDNPFKNRLSKFFKNFGVD 205
Qy	103 VLEIPK-KYQVVVISFSPEDDKERPRLGHHKEKLDGN---IAQSATKPIYLLPLDMPK 158
Db	206 SKLTQKNHEVNINDFN-----GLKEQIKKANLMTITLGANDF--LRVLDIPK 251
Qy	159 AA-----LQGYSVIVDKFNPLL--IHPL-SNFAQSITK-PLAIT 194
Db	252 LVSLNLKGNDEIKVKFENNLQKASLEITKNLKLIALIRLNSVNSITLVSPPLPL 311
Qy	195 RASDFTAKLNQPNNO---DBLWVYL---EKFPDEALKANRLQTDAPSEKGNLVDP 246
Db	312 RLQDVIDSSFKLTNKRRLSDIILNLSAIKTYD---LNNNI-----NYIDA 356
Qy	247 FVYSFIRNPQOKEWASDLNODQTVRLYLRTEPSPOAKTILKDYKYKQETFLSSIDLKA 306
Db	357 F-----DENDWKENSFNFAKAIFDIHPTELG-----YK--RMAQDIFLKM 394
Qy	307 SNGTSLFANE-----NDLKQDLVDLLDVSDYFGGOSETTISNSQVKPVPASERSLKDR 360
Db	395 SLGQN-YAQCKTASTANAIYDAWD-----SEYFNKQSD-----SFKQK 431
Qy	361 VKFKDQOKPRIEK-----FSLYEYDALSFYSQQLVSKPNSIKDLVNATLARNLRF 413
Db	432 IGFANNTQDLVEKTLGSSKGAAPWAKTALENDEELRKFSFNKFTSTIVKDWLISNKR 491
Qy	414 SLGKYNFLFDLLAS-HLDYFLVSKAIKQ-----SSITKKULFIELPI----- 455
Db	492 FGSFVNLITSNLTSLGLDKNGYISQFLSKKDAONTNNLYKLSALASTPYIDWISGIQN 551
Qy	456 ---KISL-KSSITGLDQEPNIKTLPEKVTPKLONFRDVEI-EKAFGLLYPGV-NEELBQA 509
Db	552 DLNLDJLOKNNIIGQNTISKILYD---VFRKNIINDQNLYELIKGMFHSQFNENKSOI 608
Qy	510 RKAQASPEKSKKGLKE--FSQCKEENSKAINNQOGLBEDDNIITERLPENSPIQ---- 563
Db	609 KSLIRALLSDLVSSHKLAEIIFGEFATDNLK--NNNEFLKNLLIIONKLTQTQKSLDNIIN 666
Qy	564 -----YQOENAGLGASDPKPYMTKVQNGORYIYAKSQIOELTK--AKDYTKLAKLLSNR 615
Db	667 VFVDALFONDNEYLIQGPLSNVLAKIVKTNKALIT-TRIDNFDNDFKQQTTLDA----- 720

Db	1685	-----IPKISDFAS-----LL-----DLFKYNEISEQEWKSA-----ITFFIDYLPIN 1722
Qy	1555	N-PNOK-LPENYLNVLN---QFWKVTLYNSDFITNLF-----VEPEGSDRGSGTKLKOV 1605
Db	1723	NLINSKILPTNSLLFVRNISISKTKPELNKNIENIILDLFKALSQKAKTFDNNNTLTAKNI 1782
Qy	1606	IQQVNNNYADWGSAYLTFWYDKNIITNQPNVITANIADV--FIKDVKELEDNTK 1658
Db	1783	IEFTFNK-----LLTSNNIKNFINN-----LLSNLSPIRSFLKSI-NVQDKSK 1824

Search completed: October 28, 2005, 21:08:03
Job time : 132 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2005, 21:06:45 ; Search time 224 Seconds
(without alignments)
3506.709 Million cell updates/sec

Title: US-10-607-631-8
Perfect score: 9732
Sequence: 1 MNKKSTLLATAAAIIGST.....TNNAFNNVKEFNISKIVE 1879

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1865214 seqs, 418043040 residues
Total number of hits satisfying chosen parameters: 1865214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9732	100.0	1879	15	US-10-607-631-8
2	9435.5	97.0	1878	15	US-10-607-631-20
3	475	4.9	1032	15	US-10-607-631-4
4	415	4.3	1001	15	US-10-607-631-10
5	370.5	4.8	1939	17	US-10-732-923-3340
6	350	3.6	2910	17	US-10-732-923-3342
7	346	3.6	1985	17	US-10-732-923-3351
8	325	3.3	4688	15	US-10-282-122A-76865
9	305.5	3.1	1875	15	US-10-369-493-22285
10	305.5	3.1	1875	17	US-10-732-923-3334
11	305.5	3.1	1875	17	US-10-732-923-3335

12	305.5	3.1	2503	17	US-10-828-985A-11	Sequence 11, Appl
13	305.5	3.1	2543	17	US-10-828-985A-9	Sequence 9, Appl
14	305.5	3.1	2568	17	US-10-828-985A-7	Sequence 7, Appl
15	303.5	3.1	1441	17	US-10-732-923-3352	Sequence 3352, Ap
16	303	3.1	2633	18	US-10-450-763-36864	Sequence 36864, A
17	301	3.1	2663	16	US-10-723-860-749	Sequence 749, App
18	300	3.1	1527	17	US-10-732-923-3354	Sequence 3354, Ap
19	299	3.1	2017	18	US-10-450-763-36660	Sequence 36660, A
20	297	3.1	1404	17	US-10-732-923-3304	Sequence 3304, Ap
21	295	3.0	1602	17	US-10-954-924-7	Sequence 7, Appl
22	295	3.0	1621	17	US-10-954-924-5	Sequence 5, Appl
23	295	3.0	1639	14	US-10-087-464-10	Sequence 10, Appl
24	295	3.0	1639	17	US-10-954-924-3	Sequence 3, Appl
25	291	3.0	5005	15	US-10-282-122A-76871	Sequence 76871, A
26	290.5	3.0	1679	15	US-10-369-493-22080	Sequence 22080, A
27	288	3.0	1948	14	US-10-032-585-7611	Sequence 7611, Ap
28	286.5	2.9	1979	20	US-10-205-823-419	Sequence 419, App
29	286.5	2.9	1979	20	US-11-051-454-419	Sequence 8087, Ap
30	283.5	2.9	2384	15	US-10-335-977-8087	Sequence 73, Appl
31	281	2.9	1805	10	US-09-820-843A-73	Sequence 63513, A
32	281	2.9	1805	15	US-10-282-122A-63513	Sequence 3303, Ap
33	281	2.9	1805	17	US-10-732-923-3303	Sequence 64245, A
34	280.5	2.9	1818	15	US-10-282-122A-64245	Sequence 3306, Ap
35	280.5	2.9	1818	17	US-10-732-923-3306	Sequence 8088, Ap
36	280.5	2.9	2440	15	US-10-335-977-8088	Sequence 3353, Ap
37	279.5	2.9	1478	17	US-10-732-923-3353	Sequence 23, Appl
38	275.5	2.8	10203	16	US-10-661-809-23	Sequence 4098, Ap
39	275.5	2.8	10203	18	US-10-724-972A-4098	Sequence 8312, Ap
40	274.5	2.8	1997	17	US-10-732-923-8312	Sequence 10221, A
41	273.5	2.8	1690	20	US-11-097-143-10224	Sequence 10311, A
42	273.5	2.8	1690	20	US-11-097-143-10311	Sequence 254, App
43	271	2.8	3225	16	US-10-408-765A-254	Sequence 70580, A
44	270	2.8	6641	15	US-10-282-122A-70580	Sequence 17015, A
45	267	2.7	2237	17	US-10-732-923-17015	

ALIGNMENTS

RESULT 1
US-10-607-631-8
; Sequence 8, Application US/10607631
; Publication No. US20040091901A1
; GENERAL INFORMATION:
; APPLICANT: Minion, F. Chris
; APPLICANT: Djordjevic, Steven P.
; TITLE OF INVENTION: Immunogenic Mycoplasma Hyopneumoniae
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 08411/035001
; CURRENT APPLICATION NUMBER: US/10/607,631
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392,632
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1879
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
US-10-607-631-8

Query Match 100.0% Score 9732; DB 15; Length 1879;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1879; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MNKKSTLLATAAAIIGSTVFGTVGLASKVKYGVNPTQGVISQLGLIDSVAFKPSIA	60
Db	1	MNKKSTLLATAAAIIGSTVFGTVGLASKVKYGVNPTQGVISQLGLIDSVAFKPSIA	60
Qy	61	NFTSDYQSVKALLNGKTFDPKSFPTDFVSKFDELTNNGRTVLEIPKKYQWVISEFSPE	120
Db	61	NFTSDYQSVKALLNGKTFDPKSFPTDFVSKFDELTNNGRTVLEIPKKYQWVISEFSPE	120

QY 1 MNKKSTLLATAAAIIGSTVGTWGLASVKYRGVNPQTQGVISQGLIDSVAFKPSIA 60
 DB 1 MNKKSTLLATAAAIIGSTVGTWGLASVKYRGVNPQTQGVISQGLIDSVAFKPSIA 60
 QY 61 NPTSDYOSVKALLNGKTFDPKSSEFTDFVSKEDPLTNGRTVLIEIPKKYQVVISSEPS 120
 DB 61 NPTSDYOSVKALLNGKTFDPKSSEFTDFVSKEDPLTNGRTVLIEIPKKYQVVISSEPS 120
 QY 121 DDKERFRGLFHLKEKLEJEDGNIAQSAATKFIYLLPLDMPKAALGOYSYIVDKNFNNLIHPL 180
 DB 121 DDKERFRGLFHLKEKLEJEDGNIAQSAATKFIYLLPLDMPKAALGOYSYIVDKNFNNLIHPL 180
 QY 181 SNFSAQSIKPLALTRSSDFIAKLNQFNQDDELWYLEKFPDLEALKANIRLOTADPSFEK 240
 DB 181 SNFSAQSIKPLALTRSSDFIAKLNQFNQDDELWYLEKFPDLEALKANIRLOTADPSFEK 240
 QY 241 GNLVDPFVYSIRPQNKQEWASDLNODOKTVRLYLRTSPQAKTILKDYKDETFLS 300
 DB 241 GNLVDPFVYSIRPQNKQEWASDLNODOKTVRLYLRTSPQAKTILKDYKDETFLS 300
 QY 301 SIDLKASNGTSLFANENDLKQDLDVLDVSDYFGQSEITITSQVKPVPASERSLKDR 360
 DB 301 SIDLKASNGTSLFANENDLKQDLDVLDVSDYFGQSEITITSQVKPVPASERSLKDR 360
 QY 361 VKFKDQKQPRIKPSLYEYDALFYSQLQBELVSKPNSIKOLVNAATLARNLRFSLGKYNF 420
 DB 361 VKFKDQKQPRIKPSLYEYDALFYSQLQBELVSKPNSIKOLVNAATLARNLRFSLGKYNF 420
 QY 421 LPDDLASHLDYFLVSKAKIKQSSITTKLFELPIKISLKSILGDOEPNITLPEKEVT 480
 DB 421 LPDDLASHLDYFLVSKAKIKQSSITTKLFELPIKISLKSILGDOEPNITLPEKEVT 480
 QY 481 FKLDNFRDVEIEKAFGLLYPGVNELEQAKAQRASFEKSKGLKEFSSQCKEENSKAI 540
 DB 481 FKLDNFRDVEIEKAFGLLYPGVNELEQAKAQRASFEKSKGLKEFSSQCKEENSKAI 540
 QY 541 NNQGLEEDDNIATERLPENSPYQOQENAGLSPDKPQYMIKDVONQYIYLAQIQBELI 600
 DB 541 NNQGLEEDDNIATERLPENSPYQOQENAGLSPDKPQYMIKDVONQYIYLAQIQBELI 600
 QY 601 KAKDYTKLAKLSNRHTYNI SLRKEOLFVNPRIPSSRDIEKAFVLDTKKNKYWIY 660
 DB 601 KAKDYTKLAKLSNRHTYNI SLRKEOLFVNPRIPSSRDIEKAFVLDTKKNKYWIY 660
 QY 661 SSASPVFQNKSLFGYRYLLGLDPKQTIHELVLKQKAGLQFEGYENLPSDFNLEDLN 720
 DB 661 SSASPVFQNKSLFGYRYLLGLDPKQTIHELVLKQKAGLQFEGYENLPSDFNLEDLN 720
 QY 721 IRIKTPFSQKDNFKLSLLDFNYYDGEIKAPEFGLPLFLPKELRRNSSGGGQNSNP 780
 DB 721 IRIKTPFSQKDNFKLSLLDFNYYDGEIKAPEFGLPLFLPKELRRNSSGGGQNSNP 780
 QY 781 WQEIIISQFQKQNLNQDLAQFSTKIWEKIIIGDSENFQNNRLOYLKLDLQESWINKT 840
 DB 781 WQEIIISQFQKQNLNQDLAQFSTKIWEKIIIGDSENFQNNRLOYLKLDLQESWINKT 840
 QY 841 RNLNWTYLLGDLKVKPKNNLEAKFRQISNLQELLTAFYTSAAALNNMNYQDSGAKSTI 900
 DB 841 RNLNWTYLLGDLKVKPKNNLEAKFRQISNLQELLTAFYTSAAALNNMNYQDSGAKSTI 900
 QY 901 IFEETAEALDPKVKEGADVYQLKFHYAIGPDNAGKFNQEVIRSSRTIYLTSGSKSL 960
 DB 901 IFEETAEALDPKVKEGADVYQLKFHYAIGPDNAGKFNQEVIRSSRTIYLTSGSKSL 960
 QY 961 EADTIDQLNQAENAPLQSGFVLDTEREGVFKLATSIAVHOKKEKTLPLKLNNDGYT 1020
 DB 961 EADTIDQLNQAENAPLQSGFVLDTEREGVFKLATSIAVHOKKEKTLPLKLNNDGYT 1020
 QY 1021 LIHDLKLPVPIQOISSPEKWFEGKLNQGSQNVNVTFGSIIIESPYFSTNFQEDADL 1080
 DB 1021 LIHDLKLPVPIQOISSPEKWFEGKLNQGSQNVNVTFGSIIIESPYFSTNFQEDADL 1080

QY 1081 DODGDDSRQGNLSLNDQAGLLKQKLAJLLGNQFIQYQOONKEIEFEIINVEKYSLS 1140
 DB 1081 DODGDDSRQGNLSLNDQAGLLKQKLAJLLGNQFIQYQOONKEIEFEIINVEKYSLS 1140
 QY 1141 FRVEFKLAKTLEDNGKTIIRVLSDETMSLIVNTTIEKTPKMSAVPEVFTKWEQYDPRTP 1200
 DB 1141 FRVEFKLAKTLEDNGKTIIRVLSDETMSLIVNTTIEKAPMSAAPEVFTKWEQYDPRTP 1200
 QY 1201 LAAKTFVLKFKDOIIPVDGSGNISDKWLASIPLVIHQOMLRLSPVVKTRTELGLKTEQQQ 1260
 DB 1201 LAAKTFVLKFKDOIIPVDASGNSISDKWLASIPLVIHQOMLRLSPVVKTRTELGLKTEQQQ 1260
 QY 1261 QOQQOQQOQOQKAVRKEBELETYNPKDSEFNILNPLTKAHLRTLSLNVNNDNYKIEDL 1320
 DB 1261 QOQQOQOQOQOQKAVRKEBELETYNPKDSEFNILNPLTKAHLRTLSLNVNNDNYKIEDL 1320
 QY 1321 KVIKNBAGDHLAFSLRANNIKRLMNTPTTFADYNPFYFYNEDWRSIDKYLNNKGNVSS 1379
 DB 1321 KVIKNBAGDHLAFSLRANNIKRLMNTPTTFADYNPFYFYNEDWRSIDKYLNNKGNVSS 1379
 QY 1379 KVIKNBAGDHLAFSLRANNIKRLMNTPTTFADYNPFYFYNEDWRSIDKYLNNKGNVSS 1379
 DB 1379 KVIKNBAGDHLAFSLRANNIKRLMNTPTTFADYNPFYFYNEDWRSIDKYLNNKGNVSS 1379
 QY 1380 QOQQOQOQOQOQOQKAVRKEBELETYNPKDSEFNILNPLTKAHLRTLSLNVNNDNYKIEDL 1435
 DB 1380 QOQQOQOQOQOQOQKAVRKEBELETYNPKDSEFNILNPLTKAHLRTLSLNVNNDNYKIEDL 1435
 QY 1435 LVERSIGVPWSTGLDGYIGSEOTKDGTSSSSQKGFQDQFIQALGLKNTYHKGKLSIR 1495
 DB 1435 LVERSIGVPWSTGLDGYIGSEOTKDGTSSSSQKGFQDQFIQALGLKNTYHKGKLSIR 1495
 QY 1495 IFDPGNSLAKIKDASNKKGEKLLKSYDLPKNYLNEYEKSPKIAKGTWNIHPDOKEYPN 1555
 DB 1495 IFDPGNSLAKIKDASNKKGEKLLKSYDLPKNYLNEYEKSPKIAKGTWNIHPDOKEYPN 1555
 QY 1555 PNQKLPENYLNLVNLQPKVTLVNSSDFIITNLVPEPGSDRGSGTKLQKQVIOKOVNNYA 1615
 DB 1555 PNQKLPENYLNLVNLQPKVTLVNSSDFIITNLVPEPGSDRGSGTKLQKQVIOKOVNNYA 1615
 QY 1615 DWGSAYLTFWYDKNIIITNQPNVITANIADVFIDKVKLEEDNTKLIAPNITQWPNISGSK 1675
 DB 1615 DWGSAYLTFWYDKNIIITNQPNVITANIADVFIDKVKLEEDNTKLIAPNITQWPNISGSK 1675
 QY 1675 EKFPYKTPVFFGNMENENSNMNSQOQTPTWBKIREGFAQALQKSSFDQKTRTFVLTNNAPL 1735
 DB 1675 EKFPYKTPVFFGNMENENSNMNSQOQTPTWBKIREGFAQALQKSSFDQKTRTFVLTNNAPL 1735
 QY 1735 PLWKYGLPGLQNGFNFKTQDWRLVFQNDNQIAALRVQOEDRPEKSSDKQKWKIPKV 1795
 DB 1735 PLWKYGLPGLQNGFNFKTQDWRLVFQNDNQIAALRVQOEDRPEKSSDKQKWKIPKV 1795
 QY 1795 VIPEMFNSGNIRFVGMVQIQGNPTLMLPVINSVVIYDFYRGTDGSDNDVANLNVAPQVK 1855
 DB 1795 VIPEMFNSGNIRFVGMVQIQGNPTLMLPVINSVVIYDFYRGTDGSDNDVANLNVAPQVK 1855
 QY 1856 TIAFTNNAFNVKPEFNISKIIVE 1879
 DB 1856 TIAFTNNAFNVKPEFNISKIIVE 1879

RESULT 3
 US-10-607-631-4
 ; Sequence 4, Application US/10607631
 ; Publication No. US20040091901A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Minion, F. Chris
 ; APPLICANT: Djordjevic, Steven P.
 ; TITLE OF INVENTION: Immunogenic Mycoplasma Hyopneumoniae
 ; TITLE OF INVENTION: Polypeptides
 ; FILE REFERENCE: 08411/035001
 ; CURRENT APPLICATION NUMBER: US/10/607,631
 ; PRIOR FILING DATE: 2003-06-27
 ; PRIOR APPLICATION NUMBER: 60/392,632
 ; PRIOR FILING DATE: 2002-06-28
 ; NUMBER OF SEQ ID NOS: 42
 ; SOFTWARE: FastSeq for Windows Version 4.0

Db 1645 DEGRQAMMTTLERKLAKMESQLSETKQSAEPKSVNNVQPLGLPRKIEEN--SN 1702
Qy 1312 DP-NYKIEDLVKINEAGDHQAFSLRANNIKRLMNTPIITPADYNPFYFYNEDWRSIDKY 1370
Db 1703 SPFNPLSGLKLLKNSKSSGGFN-----PFTSPSPKHLQNDNDKR---ES 1747
Qy 1371 LNNKGNVSHQOQAGNGQSGLI---ORLNKNIKPTFTPALIALKORNTNLSYSDK 1427
Db 1748 LANKTDPPTLEPSPFNIPASRLGSSSTLSTDNDEBLTNNPAQKSSNRNVQSEED- 1806
Qy 1428 IIMIKPKYLVERSICVPWSTGLDVGISEQTKDGTSSSQOKGFDQFIOALGLKNTYH 1487
Db 1807 -----TEKKGEPE-----VRGBEIEEQTKSKRPIDEVEGLANDE-- 1843
Qy 1488 GKLGLSIRIPGNEKAKIDKASNKGBEKLKSYDLFPKNYLNBYEKKSPKI 1539
Db 1844 -----DDTENINESKKIKTED-----BEEKETDKV 1869

RESULT 10
US-10-732-923-3334
; Sequence 3334, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3334
; LENGTH: 1875
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-732-923-3334

Query Match 3.1%; Score 305.5; DB 17; Length 1875;
Best Local Similarity 19.6%; Pred. No. 3.5e-08;
Matches 340; Conservative 271; Mismatches 627; Indels 494; Gaps 75;

Qy 60 ANFTSDYOSVKALLNGK-----TF-----DPKSSSETDVSFKDFEL---TNNGR 101
Db 380 AKSSSDFIPLKKLIKERRTKEHLQNIETTFIVELEHKVPIINSFKERTDMLNELNAA 439
Qy 102 TVLEIPKQYQVVISFSPEDDKERFRLGPHLKEKLEDGNIASATKFIYLLPLDMPKAAL 161
Db 440 LLL-----HTSNEKNKAKVKELNKNQKCECENDLQTLTK---QELDLCR--- 482
Qy 162 GOYSYIVDKNFNNLIHPLSNFSAOSIKPLALTRSSDFFIAKLQ-----FNNQBSLWYL 216
Db 483 -QIYL-----LITNSVNSDKGPLR---KEEIQFIQNIQWDDSDTITESDSQKVT 530
Qy 217 EKFFDLKAMIRIQATDFSEK--GNLVDPFVTSFINPONQKQEWASDLNODQKTVRL 274
Db 531 ERLVPEPKNI---IQLOEKNAELLKVVRLNADLKLESKEKSKOSLQKIESETVNEAKAI 587
Qy 275 YLRTEFSPQAKTILDKYKDE--TFLSSIDLKASNGT--SLPANENDLKOOLDVLLDV 330
Db 588 TLKSEKNDLESRIEELQKELEELKTSVPNEDASYSNVTIKQITETKRDLESQVQDLQTRI 647
Qy 331 SDYFGQSQETIT--SNSQVKVPASERSLDRVKFKDQKQPKRIEKFSLYEYDALSFYSQL 389
Db 648 SQITRESTENMSLLNKEIQDLYDSKDSISIKLGEKSSRIEAEERPKL----- 695
Qy 390 QELVSKPNSIKDLVNTATLARNRFLSGKGNFYFLDDASHLDYFVFLYSKAKIKQSSITKXL 449
Db 696 -----LSNTLDITKAEN-----DQLRKFDY---LQNTILKQDSKTHET 731
Qy 450 FIELPIKISIKSILQDQEPNIKTLPEKEVTFKLDNFRDVEIEK-----APG 496

Db 732 LNEY-VSKSKLSIVETELNLKE--EQKLHVHLEKNLKOELNKLSPKSDSLRMVTLQ 788
Qy 497 LLYPGVNEELEQARKA-QRASPEKESKGLKEFSQOKEENSAINNOEGLEDDNITER 555
Db 789 TLOKEREDLEETFKSCQKKIDELBALSELKETSOKDHHIKOL-----BEDN----- 837
Qy 556 LPENSPIOQOENAGLSPDKPMIKDVONQYRYLAKSQ--IQELIKAKDYTKLAKLLS 613
Db 838 ---NSNIWQ-QNIEALKKDYESVITSVDKQTDIEKLYKVKSLKEIEEDI----- 888
Qy 614 NRHTYNI SLKQLFDVNPRISSRDIEKAKFVLDK--TEKNKYWQIYSSASVPFQNKW 671
Db 889 RLHTYN---MDETINDDSL---KELEKSKINLTDAYSQIKYKDYETTSQLOOTN 941
Qy 672 SLFGYRYLLGLDP-----KQTIHELKVLQKAGLOFEGYENLPSPFNLEDLKNIR 722
Db 942 S-----KLDESFXDFTNQIKNLTDEKTSLEDKISLLEKEQMFNLNLEDLQ----- 986
Qy 723 IKTPLFSQDNFK--LSLLDFNNYDGEIKAPFGLPLFLPKELRRNSSNGSGSQNSNP 780
Db 987 -KKGMEKADFKKRSILQNNNKEAVEKSEYKSLKIQNDLDQOTIVANTAQNN--- 1042
Qy 781 WQEI-----ISQKDO-----NLSNODQLAQFSTKIWEKIIGENEFQ 820
Db 1043 YEOELQKHAVSKTISELRQLHTYKQGVKTLNLS-RDOL-----EN----- 1083
Qy 821 NNRLOYKLLADQESMINKTRDNLWYLYGDKLVKPKNNLEAKFROISNLQELL---TA 877
Db 1084 -----ALKENKSSWSOKESLL-----EQDLS-----NSRIELSSQNKLLOIQ 1125
Qy 878 FYTSAALSNWNYQDSGAKSTIIF-----EEIABDPKV----- 912
Db 1126 IYTAADKEVN-NSTNGPGLNILLTRERBILDTKVTVAEARDAKMLRQKISLMDVELQD 1184
Qy 913 -KEKGADVYQLKPHYAI--GFDDNAGKFNQ-EVIRSSRTI--YLTSGSKSLEADT-I 965
Db 1185 ARTKLDNRSVEKENHSIIQOHDDIMEKLANLNLRESNITLRNELENNNNKKLOSEL 1244
Qy 966 DOLNOAVKNAPLGL-----QSPYLDTERFGVFOKLATSLAVOHQKKEKTLPK 1012
Db 1245 DKLQNV--APIESALTALKYSMOEKOELKLAKEVHRWKRSODILEKEHQSSSYE 1302
Qy 1013 KLNNDGVTLIHD-----KLKPVIPQISSSPKDFEGKLNQNGSQNV 1056
Db 1303 KLESEIENLEKENKRGAEAEKFNRLRQAEERLKT-----KLSQDSLTRQV 1354
Qy 1057 N-VSTFGSIIESPYSFNFQDADLDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 1096
Db 1355 NSLRDAXNVLENSLEAN---ARIEELQNAKVAQGNQLEAIRKLOBDKASRELOAK 1410
Qy 1097 -----NOEAGLLK-----OKLAITLLGNO----- 1114
Db 1411 LEBSTTSVESTINGLNEEITTLKEEIEKQRIQOOLQATSANEQNDLSNIVESMKKSPEE 1470
Qy 1115 -FIQVYQNDKEIBFEI-----IHVEKV-----SELSFRVEFKLAKTLDENGK 1156
Db 1471 DKIFKEKTQEVNEKILEAQERLNQPSNIMNEEIKKWESEHEQEVQKIREAEALKK 1530
Qy 1157 TIRVLSDBTMSLIVNTTIE-----KTPENSAPVEVPDFTKWVQYDPTPLA 1202
Db 1531 RIRLPTKEINKIIRKEEKEFEKVEEBERIKSMEQSGEIDVVLRKOLEAKVOEKQKE 1590
Qy 1203 AKTKFVLKFDQI-PVDGSGNIS---DKMLASIPLVHQOMLRLSPVVKTIRELGLKTE 1257
Db 1591 LENEYKGLQBELKXDPVPHSSHISODDERDKLRAETESLRREEFNNELOAIKK-----KSF 1644
Qy 1258 OQ 1311
Db 1645 DEGQOQAMKTYLLERKLAKMESQLSETKQSAESPSPKSVNNVQPLGLPKIEEN--SN 1702
Qy 1312 DP-NYKIEDLVKINEAGDHQAFSLRANNIKRLMNTPIITPADYNPFYFYNEDWRSIDKY 1370
Db 1703 SPFNPLSGLKLLKNSKSSGGFN-----PFTSPSPKHLQNDNDKR---ES 1747

Db 1842 --E I E N L N A Q E L H E N L E E M K S V M K E R D N L R R V E E T K L - E R D Q L K E S I Q E T K A R D L E I 1897
 Qy 1306 S N L V N D N P Y K I E D L K V I K N E A G D H - Q L A P S L R A N N I K R L M W T P I T F A D Y N P F F Y N E D W 1364
 Db 1898 Q - - - - - Q E L K T A R M L S K E H E T V D K L R - - - E K I S E K T I Q I S D I Q - - - - - 1933
 Qy 1365 R S D K Y L N N K G N V S S H Q Q A A G N O G S G L I Q R L N K N I K P E T F T P A L I A L K D R N T W L S N Y 1424
 Db 1934 K D J D K - - - S D E L Q K D R Q - - - - - N H Q V K P E K - - - - - R L L S D G Q H L M E S L 1971
 Qy 1425 S D K I I M I K P Y L V E R S I G V P W S T G L D G Y I G S E Q T K D G T S S S Q O Q K F D D O F I Q A L G I - - - 1481
 Db 1972 R E K C S R I K E - L L K R - - - - - Y S E M D D H Y - - - - - N H Q V K P E K - - - - - R L L S D G Q H L M E S L 2003
 Qy 1482 K N T E Y H G - - - - - K L G L S I R I F D P G N E L A K I K D A S N K K G E E K L L K S Y D L F K N Y L N E Y E K K S 1536
 Db 2004 K E I F E H R I M K K L K Y L V S Y - - - - - V T K I K E E - - - - - Q H E C I N K F E M - D F I D E V E K Q 2048
 Qy 1537 P K I A G W T N I T H P Q K E Y P N P N O K L P E N Y L N L V N Q P M K V T L - - - - - Y N S S D F - - - - - 1583
 Db 2049 E L L I K - - - - - I Q H L Q D C D V S R E L R D - - - - - L K L N Q M D L H I E I L K D F S E S E F S I K T E F 2100
 Qy 1584 - - - I T N - - - - - L F V P E G S D R G S T K L Q V I Q K - - - - - Q V N N Y A D W G S A Y I T F W Y D K N 1629
 Db 2101 Q Q V L S N R K E M T Q F L E E M L N T R F D I E K L K N G I Q K E N D R I C Q V N N - - - - - F F N R I 2149
 Qy 1630 I - T N Q N V I - - - - - T A N T A D V I K D V K E L - E D N Y K L I A P N I T Q W N P I N G S K E K E Y P T V F 1684
 Db 2150 I A I N N E T E F E E R S A T I S K E W Q D L K S L K E K N E K L F K N Y Q T L K T S L A S G A Q - - - - - V N P T - - 2204
 Qy 1685 F G N W E N E N S M N S O A Q T P T W E K I R E 1709
 Db 2205 - - - T O D N K N P H V T S R A T O L T T E K I R E 2227

 RESULT 14
 US-10-828-985A-7
 ; Sequence 7, Application US/10828985A
 ; Publication No. US20050003402A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Armour, Christopher D
 ; APPLICANT: Castle, John C
 ; APPLICANT: Garrett-Engle, Philip W
 ; APPLICANT: Kan, Zhengyan
 ; APPLICANT: Loerch, Patrick M
 ; APPLICANT: Tsinoemas, Nicholas F
 ; TITLE OF INVENTION: Novel Isoforms of Centromere Protein E (CENPE)
 ; FILE REFERENCE: RS0210Y
 ; CURRENT APPLICATION NUMBER: US/10/828,985A
 ; CURRENT FILING DATE: 2004-04-21
 ; PRIOR APPLICATION NUMBER: US 60/464,905
 ; PRIOR FILING DATE: 2003-04-23
 ; PRIOR APPLICATION NUMBER: US 60/510,701
 ; PRIOR FILING DATE: 2003-10-10
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 7
 ; LENGTH: 2568
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-828-985A-7

 Query Match 3.1%; Score 305.5; DB 17; Length 2568;
 Best Local Similarity 19.3%; Pred.No.5.e-08;
 Matches 363; Conservative 305; Mismatches 646; Indels 571; Gaps 91;

 Qy 1 M K N K S T L L A T A A A I I G S T V F G T V G L A S K V K Y R G V N P T Q G V I S Q L G L I D S V A F K P S I A 60
 Db 763 I Q D K S E E L H I T S E K - - - D K L F S E V V H K E S R V - - - - - Q G L L E E I G - - - - - K T K D D L A 806
 Qy 61 N F T S D Y S Q V K K A L L N G K T F D P K S S E F T D V S K F D P L T N N G R T V L E I P K Y Q V I S E R S P E 120


```

; OTHER INFORMATION: KINESIN HEAVY CHAIN SIGNATURE domain identified by eMATRIX,
; OTHER INFORMATION: accession number PR00380A, p-value=1.00e-25, raw score of 14.18
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (12)...(2412)
; OTHER INFORMATION: Kinesin motor domain identified by Pfam, accession name
; OTHER INFORMATION: kinesin, E-value=4.9e-217, Pfam score of 734.4
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(2633)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-36864

Query Match      3.1%  Score 303;  DB 18;  Length 2633;
Best Local Similarity 19.5%;  Pred No. 7.9e-08;
Matches 370;  Conservative 322;  Mismatches 688;  Indels 520;  Gaps 91;

Qy 1 MNQKSTLLATAAAIIGSTVGTGVVGLASKVKYRGVNVPTQGVISQIGLSDVAFKPSIA 60
Db 763 IQDKSEELHIITSEK--DKLFSEVVHKESSV-----QGLLEEIG-----KTQDDLA 806
Qy 61 NFTSDYQSVKALLNGKTFDPKSEFTDFVSKDFLTNGRTVLEIPKQYQVVISFESPE 120
Db 807 TTQSNYKSTQDEFQNFKTLH-----MDFEQKYKMLEE----- 839
Qy 121 DDKERFRLGFLHKEKLEDGNIQAQATKFIYLL-----PLDMPKAAALQOYSIVIDKFNFL 175
Db 840 --NERM-----NQEVNLSKEAQKQFDSGLGALKTELSTYKQELQEKTRVQERLNEM 889
Qy 176 --IHLPLSNFSAOSIKPL-ALTRSSDFIA-KLQNFNNQDELWVLYLEKFPDLEALKANRL 231
Db 890 EQLKEQLENRDS---PLQTVREKTLITKQLQ--TLEEVKTLTQEKDDLKQLQESLQI 943
Qy 232 QTADFSFEKGNLVPFVYSFIRPNQONKEWASDLAQOKTVRLVLRTEFSQAATIL--- 288
Db 944 ERDQL--KSDIHT-VNMNIDTQELRNALSLKQHQETINT-LKSKISEVSRNLHME 998
Qy 289 -KDYKYKDEYFLSSI-----DLKASNGTSLFAN-----ENDL--- 319
Db 999 ENTGETKDEFOQKMGVGDGKQDLKAKTQITADVKNELIEQQRKIFSLQEKNELOOM 1058
Qy 320 -----KQDLVDVLLD-----VSDYFGQSETIT--SNSQKVPVPSASRS 356
Db 1059 LESVIAEQQLTKDLKENIEMTQENBELRLGDELKQKQEIIVAEQKHAIKKEGELSR 1118
Qy 357 LKDRVKFKQOQKPRIEKFSLEYDALSFYQLOELVSKPNSIKDLVNLATLARNLRFSLG 416
Db 1119 C-DRLAEVEEKLEKSKQLOEKQOQLLNVOEEMSEMQKINEIENLKNB--LKNKELTLE 1175
Qy 417 KYNFLDDELASHLDYVFLVSKAKIKOSSITKFLPIELPIKISLSSILGDQEPNIKTLP 476
Db 1176 HMETERLELAQKLNENVEEVKSITKRVKVLKEL-----QKSF 1213
Qy 477 KEVTFKLDNFRDVEIEKAFGLLYPGVNEBELQARAKQASFEKSKGLKEFSQOKEEN 536
Db 1214 TERDHLRGYIREIE--ATGL--QTKELKIAH-----IHLKEHQSTIDELRRSVSEK 1261
Qy 537 SKAINNOGLEEDDNITERLPENSPIOYQOENAGLGASPDYPMKDVONQRYYLAKSQI 596
Db 1262 TQAINTQDLSEKSHT---KLOEEIPVLHEBQEL-----LPNVKVKVSETQETMNEBEL 1311
Qy 597 QELIKAKDYTKLAKLSNRHTYNSLRLEQLFDVNPRIPS-SRDIEKAKFVLDKTEKNK 655
Db 1312 TEQSTTKDSTTLARIEMER-----LRNKEFQSOEBEIKSLTKERDNLTKIKEALEV-K 1364
Qy 656 YWQIYSSASPVFQNKWSLFGYRYLLGLDLPKQTHIE-LVKLGQKAGLQFEGYENLPSPFN 714
Db 1365 HDQL-----KEHIRETLAKIQESQSK-----EQSLN 1391
Qy 715 LEDLKNIRIKTPLFSQKONFKLSLLDFNNYVDGEIKAPFGLPLPLPKELARNSSNGS 774
Db 1392 MKEKDN--ETTKIVSEMEQFKPK-----DSALLRIEIM-LGLSKRLQESHQ----- 1435

```

```

; OTHER INFORMATION: KINESIN HEAVY CHAIN SIGNATURE domain identified by eMATRIX,
; OTHER INFORMATION: accession number PR00380A, p-value=1.00e-25, raw score of 14.18
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (12)...(2412)
; OTHER INFORMATION: Kinesin motor domain identified by Pfam, accession name
; OTHER INFORMATION: kinesin, E-value=4.9e-217, Pfam score of 734.4
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(2633)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-36864

Query Match      3.1%  Score 303;  DB 18;  Length 2633;
Best Local Similarity 19.5%;  Pred No. 7.9e-08;
Matches 370;  Conservative 322;  Mismatches 688;  Indels 520;  Gaps 91;

Qy 1 MNQKSTLLATAAAIIGSTVGTGVVGLASKVKYRGVNVPTQGVISQIGLSDVAFKPSIA 60
Db 763 IQDKSEELHIITSEK--DKLFSEVVHKESSV-----QGLLEEIG-----KTQDDLA 806
Qy 61 NFTSDYQSVKALLNGKTFDPKSEFTDFVSKDFLTNGRTVLEIPKQYQVVISFESPE 120
Db 807 TTQSNYKSTQDEFQNFKTLH-----MDFEQKYKMLEE----- 839
Qy 121 DDKERFRLGFLHKEKLEDGNIQAQATKFIYLL-----PLDMPKAAALQOYSIVIDKFNFL 175
Db 840 --NERM-----NQEVNLSKEAQKQFDSGLGALKTELSTYKQELQEKTRVQERLNEM 889
Qy 176 --IHLPLSNFSAOSIKPL-ALTRSSDFIA-KLQNFNNQDELWVLYLEKFPDLEALKANRL 231
Db 890 EQLKEQLENRDS---PLQTVREKTLITKQLQ--TLEEVKTLTQEKDDLKQLQESLQI 943
Qy 232 QTADFSFEKGNLVPFVYSFIRPNQONKEWASDLAQOKTVRLVLRTEFSQAATIL--- 288
Db 944 ERDQL--KSDIHT-VNMNIDTQELRNALSLKQHQETINT-LKSKISEVSRNLHME 998
Qy 289 -KDYKYKDEYFLSSI-----DLKASNGTSLFAN-----ENDL--- 319
Db 999 ENTGETKDEFOQKMGVGDGKQDLKAKTQITADVKNELIEQQRKIFSLQEKNELOOM 1058
Qy 320 -----KQDLVDVLLD-----VSDYFGQSETIT--SNSQKVPVPSASRS 356
Db 1059 LESVIAEQQLTKDLKENIEMTQENBELRLGDELKQKQEIIVAEQKHAIKKEGELSR 1118
Qy 357 LKDRVKFKQOQKPRIEKFSLEYDALSFYQLOELVSKPNSIKDLVNLATLARNLRFSLG 416
Db 1119 C-DRLAEVEEKLEKSKQLOEKQOQLLNVOEEMSEMQKINEIENLKNB--LKNKELTLE 1175
Qy 417 KYNFLDDELASHLDYVFLVSKAKIKOSSITKFLPIELPIKISLSSILGDQEPNIKTLP 476
Db 1176 HMETERLELAQKLNENVEEVKSITKRVKVLKEL-----QKSF 1213
Qy 477 KEVTFKLDNFRDVEIEKAFGLLYPGVNEBELQARAKQASFEKSKGLKEFSQOKEEN 536
Db 1214 TERDHLRGYIREIE--ATGL--QTKELKIAH-----IHLKEHQSTIDELRRSVSEK 1261
Qy 537 SKAINNOGLEEDDNITERLPENSPIOYQOENAGLGASPDYPMKDVONQRYYLAKSQI 596
Db 1262 TQAINTQDLSEKSHT---KLOEEIPVLHEBQEL-----LPNVKVKVSETQETMNEBEL 1311
Qy 597 QELIKAKDYTKLAKLSNRHTYNSLRLEQLFDVNPRIPS-SRDIEKAKFVLDKTEKNK 655
Db 1312 TEQSTTKDSTTLARIEMER-----LRNKEFQSOEBEIKSLTKERDNLTKIKEALEV-K 1364
Qy 656 YWQIYSSASPVFQNKWSLFGYRYLLGLDLPKQTHIE-LVKLGQKAGLQFEGYENLPSPFN 714
Db 1365 HDQL-----KEHIRETLAKIQESQSK-----EQSLN 1391
Qy 715 LEDLKNIRIKTPLFSQKONFKLSLLDFNNYVDGEIKAPFGLPLPLPKELARNSSNGS 774
Db 1392 MKEKDN--ETTKIVSEMEQFKPK-----DSALLRIEIM-LGLSKRLQESHQ----- 1435

```

```

; OTHER INFORMATION: KINESIN HEAVY CHAIN SIGNATURE domain identified by eMATRIX,
; OTHER INFORMATION: accession number PR00380A, p-value=1.00e-25, raw score of 14.18
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (12)...(2412)
; OTHER INFORMATION: Kinesin motor domain identified by Pfam, accession name
; OTHER INFORMATION: kinesin, E-value=4.9e-217, Pfam score of 734.4
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(2633)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-36864

Query Match      3.1%  Score 303;  DB 18;  Length 2633;
Best Local Similarity 19.5%;  Pred No. 7.9e-08;
Matches 370;  Conservative 322;  Mismatches 688;  Indels 520;  Gaps 91;

Qy 1 MNQKSTLLATAAAIIGSTVGTGVVGLASKVKYRGVNVPTQGVISQIGLSDVAFKPSIA 60
Db 763 IQDKSEELHIITSEK--DKLFSEVVHKESSV-----QGLLEEIG-----KTQDDLA 806
Qy 61 NFTSDYQSVKALLNGKTFDPKSEFTDFVSKDFLTNGRTVLEIPKQYQVVISFESPE 120
Db 807 TTQSNYKSTQDEFQNFKTLH-----MDFEQKYKMLEE----- 839
Qy 121 DDKERFRLGFLHKEKLEDGNIQAQATKFIYLL-----PLDMPKAAALQOYSIVIDKFNFL 175
Db 840 --NERM-----NQEVNLSKEAQKQFDSGLGALKTELSTYKQELQEKTRVQERLNEM 889
Qy 176 --IHLPLSNFSAOSIKPL-ALTRSSDFIA-KLQNFNNQDELWVLYLEKFPDLEALKANRL 231
Db 890 EQLKEQLENRDS---PLQTVREKTLITKQLQ--TLEEVKTLTQEKDDLKQLQESLQI 943
Qy 232 QTADFSFEKGNLVPFVYSFIRPNQONKEWASDLAQOKTVRLVLRTEFSQAATIL--- 288
Db 944 ERDQL--KSDIHT-VNMNIDTQELRNALSLKQHQETINT-LKSKISEVSRNLHME 998
Qy 289 -KDYKYKDEYFLSSI-----DLKASNGTSLFAN-----ENDL--- 319
Db 999 ENTGETKDEFOQKMGVGDGKQDLKAKTQITADVKNELIEQQRKIFSLQEKNELOOM 1058
Qy 320 -----KQDLVDVLLD-----VSDYFGQSETIT--SNSQKVPVPSASRS 356
Db 1059 LESVIAEQQLTKDLKENIEMTQENBELRLGDELKQKQEIIVAEQKHAIKKEGELSR 1118
Qy 357 LKDRVKFKQOQKPRIEKFSLEYDALSFYQLOELVSKPNSIKDLVNLATLARNLRFSLG 416
Db 1119 C-DRLAEVEEKLEKSKQLOEKQOQLLNVOEEMSEMQKINEIENLKNB--LKNKELTLE 1175
Qy 417 KYNFLDDELASHLDYVFLVSKAKIKOSSITKFLPIELPIKISLSSILGDQEPNIKTLP 476
Db 1176 HMETERLELAQKLNENVEEVKSITKRVKVLKEL-----QKSF 1213
Qy 477 KEVTFKLDNFRDVEIEKAFGLLYPGVNEBELQARAKQASFEKSKGLKEFSQOKEEN 536
Db 1214 TERDHLRGYIREIE--ATGL--QTKELKIAH-----IHLKEHQSTIDELRRSVSEK 1261
Qy 537 SKAINNOGLEEDDNITERLPENSPIOYQOENAGLGASPDYPMKDVONQRYYLAKSQI 596
Db 1262 TQAINTQDLSEKSHT---KLOEEIPVLHEBQEL-----LPNVKVKVSETQETMNEBEL 1311
Qy 597 QELIKAKDYTKLAKLSNRHTYNSLRLEQLFDVNPRIPS-SRDIEKAKFVLDKTEKNK 655
Db 1312 TEQSTTKDSTTLARIEMER-----LRNKEFQSOEBEIKSLTKERDNLTKIKEALEV-K 1364
Qy 656 YWQIYSSASPVFQNKWSLFGYRYLLGLDLPKQTHIE-LVKLGQKAGLQFEGYENLPSPFN 714
Db 1365 HDQL-----KEHIRETLAKIQESQSK-----EQSLN 1391
Qy 715 LEDLKNIRIKTPLFSQKONFKLSLLDFNNYVDGEIKAPFGLPLPLPKELARNSSNGS 774
Db 1392 MKEKDN--ETTKIVSEMEQFKPK-----DSALLRIEIM-LGLSKRLQESHQ----- 1435

```


Db	1312	TEQSTTKDSTTLARIEMER-----LRNKEFQESQBEIKSLTKYERDNLTKIKEALEV-K	1364
Qy	656	YQWYSSASPVFONKWSLFGYRYLLGLDPKQTHIE-LVKGQKAGIQFEGYNLPBDFN	714
Db	1365	HDQL-----KEHIRETLAKIQESQSKQ-----EQSLN	1391
Qy	715	LEDLKNRIKTPFSOKDNPKLSLLDFNNYDGEIKAPFGLPLFLPKELRNSNSGGS	774
Db	1392	MKEKN--ETTKIVSEMQFQPK-----DSALLRIEIEIM-LGLSKRLQESH	1435
Qy	775	QNSNSPQOEIISQFQDNQDQAFSTKIWE---KIIGDNEFDQNNRLQYKLLK	831
Db	1436	-----EWSVAKEDDLQRLQEVLOSQDKENIKEIVAKHLETEBELKVAHCLKE	1488
Qy	832	LOESWINKTRNL-----YWTYLGDKLVKPKNNLEAK-----FQIENLQELL	875
Db	1489	-----QEEINELRVNLSEKETEISTIQLEAINDKLQNKIQEIVEKEQNLKIQISEVQEN	1547
Qy	876	TAF-----YTSALSNNWYQDSGAKSTIIFETIAELDPKVKKEVGADV	920
Db	1548	NELQKQFKEHRKAKDSALOSIESKMLELTNRLQESQEEIQIMIKEKEM-KRVQEAQIER	1606
Qy	921	YOLKPHYAIGFDNAGKQNOEVIKSSRTI-YLTKSGSKLEAD--TIDQLNQAQNAPL	977
Db	1607	DQK-----ENTKEIVAKKESQEKYQFLKMTAVNETQEKMCIEHLKEQFETQKL	1658
Qy	978	GIQSYFLDTERGVFQKLATSIAVOH-----KQEKTLPKLNDGTYTLIHDKLLK	1028
Db	1659	NLEN-IETENIRLQILHENLEEMRSVTKERDDLSVEETL--KVERD--OLKENLRE	1711
Qy	1029	PWIPQISSPEKDWEGKLNQNGQSNVNSVFGSIIESPYSTFQFQDADLDQDQ---	1084
Db	1712	TITROLEKQEBELKIVHMLKHEH---QETIDKURGIVSEKTNIEISNQKDLHSNDALKAQ	1768
Qy	1085	-----ODDSQGNNSLDNQB-----AGLLKQKLAILL-----GNQFIQYQONDKEI	1126
Db	1769	DLKIQEELIAHMLHKEQETIDKLRGIVSEKTKLSNNQKDLNSNAKIQELKAN	1828
Qy	1127	EFEEIINV-----EKVSELSFRVEFKLAKTLBNGKTIIRVLSDETWSLIVNTTIEKTP	1178
Db	1829	EHLQITLKQDVNETQKVSEME-----QLKQIKQDSLTLSKLEIENLNL-AQELHENLE	1882
Qy	1179	ENSAPVEVDT-KWVEQ-----YDPTPLAATKFKVLKQDQIPVDG	1219
Db	1883	EMKSVMKERDNLRRVEETLKLERDQKESQETKARDLEIQOELTKARMLSEKHEKTV	1941
Qy	1220	SGNISDKLASIPLVTHQOMRLSPVVKTIIRGLKTEQOQOQOQOQOQOQOQKAVRKE	1279
Db	1942	-----KLREK-----ISEKTIQISDIQK-----DLQKSKDELQKKIQELQKELQLLRVK	1986
Qy	1280	BELETYNPKDEFNINPLTKAHLRLTSLNLVNDPNYKIEDLVKINEAGDHQALAFSLRAN	1339
Db	1987	EDVNMSHKK--INEMEQLKKQF-----EPNY-----LCKEMDNFQLTQKLH--	2026
Qy	1340	NIKRLMNTPTPADNPPFYNYEDWRSIDKYLNNKGN--VSSHQOQAAGNGSGGLIQLR	1397
Db	2027	--ESLEIRIVAKE-----RDELRIKESUKMERDQFIATLREMIADRO-----	2069
Qy	1398	NKNIKPTETTPALIALKDRNNTNLSYSDKIIMIKPKYLVERSIGVPMWSTGLDGYIGSEQ	1457
Db	2070	NHQVKEK-----RLISDGOOHLMESLREKCSRIKE--LLKR-----YSEMDHY----	2112
Qy	1458	TKDGTSSSSQKQFQDQFQALGL---KNTEYHG-----KLGLSIRIFDPGNELAKIKDA	1509
Db	2113	-----ECLNRLSLDLEKETEIEFHRIMKCLKLYLSY-----VTKIKEE	2148
Qy	1510	SNKGBEKLKSYDLFPKYNLYNEKSKPKIAGKWTNIHPDQKPEYPPNOKLPENLYNLVL	1569
Db	2149	-----QHECINKPEM--DFIDEVEKQKELLIK---IQHLQOQDCVPSRELDR-----LKL	2193
Qy	1570	NQPMKVTL-----YNSDPE-----ITN-----LFVEPEGSDRSGTKLQVIOK	1608
Db	2194	NQNMDLHIEILKDFSESEPPSIKTEFQQVLSNRKEMTQFLEWLNTRFDIEKLKNGIQK	2253
Qy	1609	-----QVNNYADMGSAVLTFWYDKNI-ITNOPNVI-----TANIADVFIKDVKEL-EDNT	1657
Db	2254	ENDRICOVNN-----FFNNRIIIMWSESTEFERSATISKWEQDLKSLKEKNE	2302
Qy	1658	KLIAPNITOWPWNISGSKFKYKPTVFFGNWENENSSMNSOAOPTPTWEKIRE	1709
Db	2303	KLFPNYQTLKTSLSGAQ-----VNPT-----TQDNKNPHVTSRATQLTTEKIRE	2347
RESULT 18			
US-10-732-923-3354			
; Sequence 3354, Application US/10732923			
; Publication No. US20050108791A1			
; GENERAL INFORMATION:			
; APPLICANT: Edgerton, Michael D			
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES			
; FILE REFERENCE: 38-15(52796)C			
; CURRENT APPLICATION NUMBER: US/10/732,923			
; CURRENT FILING DATE: 2003-12-10			
; PRIOR APPLICATION NUMBER: 10/310,154			
; PRIOR FILING DATE: 2002-12-04			
; NUMBER OF SEQ ID NOS: 24149			
; SEQ ID NO 3354			
; LENGTH: 1527			
; TYPE: PRT			
; ORGANISM: Plasmodium yoelii yoelii			
; FEATURE:			
; NAME/KEY: unsure			
; LOCATION: (1)..(1527)			
; OTHER INFORMATION: unsure at all Xaa locations			
US-10-732-923-3354			
Query Match 3.1%; Score 300; DB 17; Length 1527;			
Best Local Similarity 19.4%; Pred. No. 5.3e-08;			
Matches 336; Conservative 278; Mismatches 620; Indels 500; Gaps 81;			
Qy	97	TNNGRTV-LETP-----KKYQVV-----IGEF--SPEDDKERFRLGFLHKELEDGNIQA	143
Db	113	TRSARTVAEILPGTKINTKNTIVQGGIISPYMSSNIDKNK-----NIPN	158
Qy	144	SATFPIYLLPLDMPKAAALQGYIYVDKFNNNLIITHPLSNFSAQSIKPIALTRSSDFIAKL	203
Db	159	DNKKELFAFTFIKPESE-----NEENKEBREVEP--NINKQNEK-----RHDGDIKPE	204
Qy	204	NQFNQDELWYLYLEKFFD--LEALKANIRLQATD-----SFEKGNLVDPFVYSFINP	255
Db	205	EEFTNN-----LNKISDNVLGDKNNIIPFESNNHNDKERTSLSKONL-----FNLNE	251
Qy	256	QNOKEWASDLNODQKTVRLYLRTFSPQAQTKLDYKYKDETFLSSIDLKASNGTSLPAN	315
Db	252	INKKIDSK-NDDNNNLLFSENO-----TLIPAKVNEVTEIEAS	290
Qy	316	ENDLKDQJVDLLDVSDYFGQSETITSNSQVKPVA--SERSLKORVKFKDQOKPRI	372
Db	291	RTP-----NYEINRVNPEYINKSGIHPNIQFNNPAQLPTEINIKNNINY--DNISPKI	343
Qy	373	EKFSIYE-----YDALSFYQLQELVSKPNS---IKDLVNATLARN--LRFS	414
Db	344	NSKGVIENLENKKEIKRISMDFNFEFNNPIENHINKRNTTIVRSVFPQLEYNNVENE	403
Qy	415	LGKYNFL--PDDLASHLDYVFLVSKA--KI-----KQSSI-TKKLFIELPIKIS	458
Db	404	IRKFTIAGNNYANIGRENDIINISPKVIEKINRRNTHQLFWNESSINSTDKLTIHKVTSPF	463
Qy	459	LKSSILGQBOEPNIKTLEKVTFLKDNFRDVE-IEKAFGLLYPGVNEELEQARQAR-AS	516
Db	464	LFDNINNSNTNV-----SNEIPEKGGKMEKVEKMEKM-----EKMEKMEKVEKME	510
Qy	517	FEK-EKSKKGLKEFSQ-----QKEEN-----SKAIINQGLE-EDDNI	552
Db	511	FEKMEKSEHNVSFSFMFLDKGVSVSTSNITAMDENFIILNDSNTINNNIIDAESKL	570

181 QY -----SNFSAQSIKPLATRSDFIAKLNQPNODELMWYLEKFDLEALKANRLQOTAD 235
 479 Db DIAGFEIPDINSFEOLCINYNTY---KIQPFNFHSF---ILEQ---SEYLRENINWEFID 530
 236 QY FPFKEGNLVD-----PPVYSFIRPNQKQEWASDLNODQKOTVRLYLRTBSPQAKTILK 289
 531 Db FGQDLOPTDILIEYKQPMGIIKLDEECMLPKPSDASPMELKSNKFNTH----- 580
 290 QY DYKYKDETFLSSIDLKASNGTSLFANENDLKQDLD-----VDLLD-----VSDYF--- 334
 581 Db KKPSFNKNGFIHHVAGVYENWNLQKNTDPTISESILNLLPDSQNELIADMFIND 639
 335 QY -----GGSETITSQKVPVPSASERSLKORVFKKQOQK-----PRIEK 374
 640 Db PHINRPQNGG-----GNSKLK--TASQKH-KQDLKTLMDQLESTEPHFVRCILPNLEK 690
 375 QY FSLVEYDALSFYSQLOELVSPNSIKDLVATLARNLRFSLGKY--NFLPDDLASHLDYY 432
 691 Db RANKFDKMLVGLQR-----CNGVL-EGIRITRAGVPMRMFDEF---TQRY 733
 433 QY FLVSKAKIKQSSITKFLPIELPIKISLSSILGQEPNI---KTLFEKEVTFKLDNPRDV 489
 734 Db SIICNDEL--SSPQNTCE---IILKFKVKNPEDFKVGLTKIFFKNGILGKLEIIRDL 787
 490 QY EIEKAFGLLYPGVN-----ELEQARKAQRASFEKEKSKG---LKEFSQOK- 533
 788 Db ALKNFTDLQKVRGNLRLVQLKQIKQISQAQIISRTWVTLDEIKSNPWWRLFFHVKP 847
 534 QY -----EENSKAINN--QEGLE-----EDDNITERL-PENSPLOYQOEN-----AGL 571
 848 Db LLEDKAKVLDKSKLQENIQTTLVKDKSEKUTKGLTNEKREOMNLMQDEMINTISIA 907
 572 QY GASDPKPYMKDVQNRQYVLAQSQOEL-IRAKDYTKLAKLSNRHTYINISRLKEQLFD 630
 908 Db KEKDKLQSLRLTENK---SKHRIETLEIKLQKPEKQNDLINEHE-----KLTRRSLE 958
 631 QY VNPRIPSSRDEKAKFVLDKTEKNKYQWYISASPVFQNKWSLFGYYRYLLGLDPKQTIH 690
 959 Db LH-----DKHNAKV---EBLNSLNKLHDSQAQSELSNL-----KQOLK 992
 691 QY ELVKGQAGLQFEGVENLPDFNLDELK-----NIRIKPLFSOKD----- 732
 993 Db EITKLNQAHSEKLLPKQM-HDKSIEDLNEKVSRLNENLKKLQADLNSKDVISEHGLK 1051
 733 QY -NFKLSLLDFNNYDGEIKAPE---FGLPLFLPKELRRSNSGSGSNSPWEQEI--- 785
 1052 Db QNEIMKLELCSDYERKQKDSORTIDSILQILKKD-----EDYKSYESKIEBA 1100
 786 QY -----ISQKQDNLSNODQLAQFSTKIWEKIIGDENEPDQNNRLQYKL 828
 1101 Db REKVTLKKGVBKKSQEIQDYKSEIKKLKAEIASSNTKI-----TQDQSTK-ELSS 1151
 829 QY LKDLQESW---INKTRDNLWYTLVGDKLKVKPKNLEAKFR---QISNLQELLTAFTVS 881
 1152 Db LKSNESKYVAIESAKQQL-----SKKIEYSSTEADYNLQTELQKMKQNTYEYSTR 1204
 882 QY -AALSNMNNYQDSKAKSTI-----IFEEIAELDPKVKVKGVADVYOLKPHYAIG 930
 1205 Db ITELSTKLNOVTEAKSKEIENKQPPNPQPMEEFTHMKLVNEQT-ASLRKEKENK-K 1262
 931 QY FPDNAGKFNQVIRSSRTIYKTS-----GKSKLEADTIQOLNOAVNAPLQOSFVL 984
 1263 Db LSEELQMLKERIMWGLSTMDLTPKRRSLAIGDKSMITVTVDSPFNKEIENLKFOLQ---- 1318
 985 QY DTERFVQFKLATSIAVQHKQEKTLPKLNDGYTLIHDKLKXPVLPQISSSPKDWFE 1044
 1319 Db --OEQGNFOR-ABENYAI---ELQKLNKLTLTRGLNTNTD-----YE 1354
 1045 QY GKLNQNGQSONVNVSTFGSIIHSPYSTNFQBDADLDQDQD-DSRQ-----GNSLND 1097
 1355 Db KKYNDQSKRITQLETKIGRLLAN---SGGDYEDHLLDNDDDWESRSSNGSGSGSTSSS 1411

1098 QY QEAGLLKOKLAILLGN-----OFIOVYQOQNDKEIEF--EIIINVKVSELSFRV-----E 1144
 1412 Db ARNSLVKSESITAFHNRGVSQDYIYQIDITTKLSTREELANGSKTEILRLKALLRESE 1471
 1145 QY FLAKTLEDNGKT-IRVLSDETMSLIV--NTTIEKTEPMSAVPEVFDTKWVEQYDPRTPPL 1201
 1472 Db DELVQVQKQENYKTSVHDYEQDLAQLKVKGHETLLSRNKNIDINESLEIY-KKGSDEYVKKLEL 1530
 1202 QY AAKTKFVLK-FKQIIPVDSGSGNISDKWLASIPLVIHQOMLRLSPVVKVTIRELGLKTRQOQ 1260
 1531 Db AESAIAISKHBEOATKEMKESRSQ-----LLLVREELRTTQIL--IKDFRIKVENLE 1581
 1261 QY QOQOQOQOQOQKQKAVRKEBELETYNPKDEFNINL-----PLTKAHRILTSLNVLNNDPNYK 1316
 1582 Db ATIEEKHQLDANK-----HEIKQIQDKLNYHLKNFENKELNEKLKEEIKNL-NRDLDFK 1635
 1317 QY IE-DLKVIKNEAGHQIAFS---LRANNI-----KRLMNTPIITADYNPFFYYN 1361
 1636 Db TDITETLIK-ENKKQLQJEDYEDVLLVKNLQNEVEEELILOBEKLQN----- 1679
 1362 QY EDWRSIDKYLNNKNVSSHQOQAGGQSGSLQRLNKNIKPETFTFPALIALKDRNNTNL 1421
 1680 Db KIDELTNNNRQLEN-----EKLNERKIVNCTKQISGLKE----- 1714
 1422 QY SNYSDKIIMIKPKYLVVERSIGVPWSTGLDGYIGSEOTK-DGTSSSQKGFDDQFIQ--- 1477
 1715 Db -LVDELSIERDKLLKDKET-----LQNDLQTMWTKNFDAITTELQAQHGELDFLKKHL 1765
 1478 QY -----ALGKNTYEHGKGLGSLIRIFDP---GNELAKIKDAS---NKKGBEKLKSYDL- 1524
 1766 Db ENQREDSEAIKTELNOQSKMSTSPDIRQOKLRNELLVTKENFSLVTKNELNLKVSLE 1825
 1525 QY FKNYVNE-----YEKSKPIAKGWTNIHPDQKEYPNQKLPENYLNL---VLNQPKWVTL 1577
 1826 Db EKLYSNQLKYWESKVDTLKALDGNALNEKHEADKTIKNLQORSIKQLEIRVENESQLSKR 1885
 1578 QY YNSDDF 1583
 1886 Db YNDENF 1891

RESULT 28
 US-10-205-823-419
 ; Sequence 419 Application US/10205823
 ; Publication No. US20030108963A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Schlegel, Robert
 ; APPLICANT: Monahan, John E.
 ; APPLICANT: Endege, Wilson O.
 ; APPLICANT: Gannavarapu, Manjula
 ; APPLICANT: Gorbacheva, Bella
 ; APPLICANT: Hoersch, Sebastian
 ; APPLICANT: Kamatkar, Shubhangi
 ; APPLICANT: Womsey, Angela M.
 ; APPLICANT: Glatt, Karen
 ; APPLICANT: Zhao, Xumei
 ; APPLICANT: Anderson, Dustin
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
 ; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
 ; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
 ; FILE REFERENCE: MRI-044
 ; CURRENT APPLICATION NUMBER: US/10/205,823
 ; CURRENT FILING DATE: 2002-07-25
 ; PRIOR APPLICATION NUMBER: 60/307,982
 ; PRIOR FILING DATE: 2001-07-25
 ; PRIOR APPLICATION NUMBER: 60/314,356
 ; PRIOR FILING DATE: 2001-08-22
 ; PRIOR APPLICATION NUMBER: 60/325,020
 ; PRIOR FILING DATE: 2001-09-25
 ; PRIOR APPLICATION NUMBER: 60/341,746
 ; PRIOR FILING DATE: 2001-12-12
 ; PRIOR APPLICATION NUMBER: 60/362,158
 ; PRIOR FILING DATE: 2002-03-05

Query Match	2.9%	Score 286.5	DB 14	Length 1979
Best Local Similarity	19.0%	Pred. No. 4.6e-07		
Matches 373	Conservative 345	Mismatches 682	Indels 563	Gaps 99
US-10-205-823-419				
QY	22	FGTVVGLAKVYRGVNPQTQVVISQGLGIDSVAKPSIANFTSD-----VQSVKAL	73	
DB	164	FGDISSQEQIN-RLSNEVSRLESEVGHWRHIA-QTSAQOTDSDQSEICKLQNIKEL	221	
QY	74	LNGKTFDPKSEFTDFVSKFPLFN-NGRTVLEIPKKYQVVISFEPSDEDDKERFLGHL	132	
DB	222	KQN-----RSEIDHQHEMVLNQAHQCKUTEISRRHRELSDY-----EERI-----	265	
QY	133	KEKEDGNIAGSATKFIYLLPLDMPKALGOYSIVDKNFNLIHPLSNFSQAQIKPLA	192	
DB	266	-EEL- -NLLQGGSGV--IETDLSKI-----YEMQKTIQVLEKVE--STKWEQLE	312	
QY	193	LTRSSDFIAKLQFN-----ODELWVLEKFFDEALKANIRLO-----TA	234	
DB	313	-DKIKINDKLLSSAENDRILRRQEQLNV--EKQIMEECE-NLKECSKLQPSAVKQS	368	
QY	235	DFSEKGNLV-----DPF-VYSFIRNPQKQWASDLNQD-----QKTVRLYLRTFSP	282	
DB	369	DTMTEKERITAQSAVEEVPFLQALGDAENEIMRLSLSLNQDNSLAEDNLKMKRMEVLE	428	
QY	283	QAKTILDKYKQDETFLSSIDL-----ASNGTSLPANENDLKQDLVDLLDVS DY	333	
DB	429	KEKSLLS--QKBELOMSLLKLANEYEVKSTATRDLSLSELHDLRLNE-----	477	
QY	334	FGQSETITNSQVKKPVPASERSLKDRVKFKQKQKPRIKFSLYEYDALSPYSQLOELV	393	
DB	478	-----AKEQLNQSTIS-BKETLIAEIBELDRQNEATKHMILIKDQL	518	
QY	394	SKPNSIKDLVNATLARNLPSLGKYNFLD---DLASHLDYFLVSKAKIKQSSITKLF	450	
DB	519	SKQNEGDSIISKQLDNDKBEKHVQLEDKMDITKELD-----VOKEKLIQSEVA---	570	
QY	451	IELPKISLKSILGDQEPNIKTLFEKVTPLKNFRDVEIEKAFGLYPGVNEELQAR	510	
DB	571	-----LNDLH-----ITKQKLEDKVENL-----VDQLN	593	
QY	511	KAQASPEKSKGKLFESQOKBENSKAINNQ--EGLEEDN-----ITER-----	555	
DB	594	KSQSSNVSIOKENLEKEHIRQNEEELSRINELMQSLNQSDNSNPNKDTLLKEREAEVRN	653	
QY	556	LPEN-SPIYOQOENAGLQSPKPYMKDVQNYIYA-----KSQIELKADYTKLAK	610	
DB	654	LKQNLSELEQLNENLKKVAFDVK-----MENEXLVLACEDVRHQLECELAGNNQLSLEK	707	
QY	611	LLSNRHTYNISLRKEQLFDVNPRI PSSRDIEKAKFVLDTKKNKYWQIYSSASPVFQNK	670	
DB	708	-----NTIVETLMEKEGIEAELCWAKKRLLEEA-----NKYEKTTIEUSNA--RN-	751	
QY	671	WSLFGYRYRLGLDLPKQTIHE-LVKLGQKAGLQF-----EGYENLPSDF-----NLE	716	
DB	752	-----LNTSALQLEHEHLIKLQKQDMETAEELKKNIEQMDTDHKTQKOVLSSELE	801	
QY	717	DLKNIRIKTFLPSQKNFKSLDLFNYYDGEIKAPFGLPLFLPKELRN-----	767	
DB	802	BQQLQ---TQLINKKEIFIEKLRSSKQBELDK-----YSQALRKNEILRQTIEE	850	
QY	768	SSNCGGSSONSPEOEIISQFKDQ-----NLSNODOLA-----QFST-----	805	
DB	851	KDRSLGSMKEENHLOEELRLEBEQGRTPAVDPKTLDSVTELASVQSQNTIKHELEE	910	
QY	806	--KTWEKIIIGBENFDQNNRLQYKLLKDLQBSWINKTRDNLWYTLGDKLKVPKPNNLEA	863	

US-11-051-454-419
; Sequence 419, Application US/11051454
; Publication No. US20050191673A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Womsey, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/11/051,454
; CURRENT FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: US/10/205,823
; PRIOR FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 419
; LENGTH: 1979
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-051-454-419

Query Match 2.9%; Score 286.5; DB 20; Length 1979;
Best Local Similarity 19.0%; Pred. No. 4.6e-07;
Matches 373; Conservative 345; Mismatches 682; Indels 563; Gaps 99;

QY 22 FGTWGLASKYRGVNPQTQGVISQLGLDLSVAFKPSIANFTSD-----YQSVKAL 73
DB 164 FGDIISSQGEIN-RLSNVSRLESEVGHWRHIA-QTSKAQGTDSQDGEICKLQNIKEL 221
QY 74 LMGKTFDPKSSBFTDFVSKFDFLTN-NGRTVLEIPKQYQVVLSEFSPEDDKERPLGFHL 132
DB 222 KQN-----RSQEIHHQHEMSVLQNAHQOKLTEISRHRBELSDY-----EERI----- 265
QY 133 KEKLEDGNTAQSNATFIYLLPDMKPAALQGVSYIVDKNFNNLIITHPLSNFSAQSIKPLA 192
DB 266 -BELE--NLQOQGGSGV--IETDLSKI-----YEMQKTIQVQLQEKVE--STKKWEQLE 312
QY 193 LTRSSDFIAKLQPNN-----QDELWVYLEKFFDLEALKANIRLQ-----TA 234
DB 313 -DKIKDINKLSAENDRIILRREQQLNV--EKQIMEECB-NLKLECSKLQPSAVKQS 368
QY 235 DFSFEKGNLV-----DPF-VYSFIRNPQONKEWASDLNOD-----QKTVRLYLATERSP 282
DB 369 DWTTEKERILQASASVEEVRLLQQAALSDAENIMRLSSLNQDNLAEADNLKLMKRIEVL 428
QY 283 QAKTILKDKYKDETFLSSIDLK-----ASNGTSLFANENDLKOOLDVLLDVS DY 333
DB 429 KEKSLLS--QKEELQMSLLKLNNEVEVIKSTATRIDLSDELHDLRLNLE----- 477
QY 334 FGQSQSETITSNQVQKVPASERSLKDVRPKKQQKPRIEKFSLYEYDALSFSYQLQELV 393
DB 478 -----AKEQELNQIS-EKETFLAIEELDRQOEATKMHILIKQOL 518

QY 394 SKPNSIKDLVNATLARNLRFSLGKNYFLFD--DLASHLDYDFVFLVSKAKIKOSSITKKLF 450
DB 519 SKQONEGDSIIISKLQDLNDEKRVVHQLDEKWDITKELD-----VQEKLIQSEVA----- 570
QY 451 IELPIKISLKSSILGDOEPNITKLFBEKVTFFKLDNPRDVEIEKAFGLLPGVNEEELQAR 510
DB 571 -----LNDLH-----LTQKLEDKVENL-----VDQLN 593
QY 511 KAQRASFEKEKSKGLKEFSQOKEENSKAINNO--EGLEEDDN-----ITER----- 555
DB 594 KSQESNVSIQKENLELKEHIRONEEELSIRINMELMSLNQDSNSNFKDITLLKEREAEVRN 653
QY 556 LPEN-SPIOYQOENAGIGASDPKPYMIKDVQNYRYLA-----KSQIQELIKAKDYTKLAK 610
DB 654 LKQNLSELEQLNENLKKVAFDVK-----MENEKVLACEDVRHQLSECLAGNQLSLEK 707
QY 611 LLSNRHTYNISLRLKEQLFVNPRIPSSRDIEKAKFVLDKTEKNKYQWIIYSSASPQNK 670
DB 708 ----NTIVETLWKEKEIEAELCWAKRILLEA-----NKYEKTIEELSNA-RN- 751
QY 671 WSLFGYRYLLGLDPKQTIHE-LVTLGQKAGLOF-----EGVENLPSPDF-----NLE 716
DB 752 -----LNTSALQLEHEHLIKNQKQKMEIAELKKNIEQMDTHKETHKDVLSLSLE 801
QY 717 DLKNIRIKTPLFSOKNFKLSLLDFNNYDGEIKAPEFGLPLFLPKELRRN----- 767
DB 802 EQKQL-----TQLINKKEIFIEKLKXERSKLOELDK-----YSQALRKNELRQTIEE 850
QY 768 SSNSGGSONSNPWEQIIISQFKDQ-----NLSNQDQLA-----QEST----- 805
DB 851 KDRSLGSMKEENHLOEBELERREEQSRTPADPKTLDVTELASVSLNTIKEHLEE 910
QY 806 --KIWEKIIIGDENEFQNNRLOYKLLDQESWINKTRDNLWYTLGDKLVKPKNNLEA 863
DB 911 EIKHQKIIEDQNG-----SKMQ--LLQSLQEO--KKEMDE--FRYQHEQMNATHQTFLE 960
QY 864 KFRQISNLQELLTAFYT-----SAALSNNNWYQDSGAKSTI 900
DB 961 KDEEIKSLQKTIQIKTOLHEERODIQTDNEDIFQETKVQSLNTEENGSEKHDLSKAETER 1020
QY 901 IFEBIAELDPKV--EK---VGADVYQLKPHYAIGFDDNAGKFNQEV----- 942
DB 1021 LVKGIKERELEIKLLEKNISLTQIDQLS-----KDEVGKLTQIIQKDLFIQALHA 1073
QY 943 -IRSSST--IYLTKSGSKLEADTIDQ-----LNAQVKNAPLGLQSFYLDTRFQVFP 992
DB 1074 RISSTSHTDVVYL-----QOOLQAYAMEREKVPFVLNEKTR-----ENSHLKTXYHKOM 1123
QY 993 QKLATSLA--VQHKQKERTLPKLNNDGYTLIHDKLKPKVIPQISSSEPKDMFEGKLNQ 1050
DB 1124 DIVAAKEAALIKLODENKKLSTRFESSQDMFRETION--LSRIIR--EKDIEIDALSQK 1179
QY 1051 QGS-----QNVNVTFGSIIIESPYSTNFQDADLD-----QDQDSDS 1088
DB 1180 CQTLAVLQTSSTGNEAGGVNSHOFELLOBERDKLKQOVKQKMBEKKQOVMTTVQNMOHES 1239
QY 1089 ROGNNSLDNQAGLL-----KQKLAILLGNQFIOVYQONDKEI----- 1126
DB 1240 AQLOEELHQLQVLVDSDNNNSKLQVDY-TGLIQSYEQNETKLNKFGQELAQVQHSIGQL 1298
QY 1127 -----BFEIIN-----VEKVSSELSPRVFKLAKTUED 1153
DB 1299 CNTKDLLGLKLDIIISPOLSSASLLTPQSAECLRASKSEVLSESSSELLQOEELRKSLOE 1358
QY 1154 NGKTRIVLSDE-----TMSLIWNTTIEKTPEMSAPVEPDTKWVEQYDPRTPLAATKTVL 1209
DB 1359 KDATIRTLQENNRHRLSDSIAATSELRKEHBQTDSEIKQLK--EKQDVLQKLLKEKDLILI 1416
QY 1210 KFKQIIPVDGSGNISDKW-----LASIPLVIHQOMLR-----SPVVKTIIRBELG 1253
DB 1417 KAKSQDLSSENFTNKNVENELLRQAVTNLKERILLILEMDIGKLGKGENEKIVEYR--G 1474
QY 1254 LKTEQOQQOQQOQQOQQOQKKAVRKEEBELETYNPKDEFNINPLTKAHLRTLNLVNDP 1313

Db 1475 KETEQALQETNMKFSM-----MLREKEFECHSMKRALFAEQLLKEBQKGTGELNQLL 1529
Qy 1314 N--YKIEDLVKINEAGDHOLAFSLRANNTIKRLMNTPTITADYNPPFYNNEDWRSIDKYL 1371
Db 1530 NAVKSMQKTVVFOQERD-QVMLAKO---KQMENTAL-----QNEVQRLRDK-- 1573
Qy 1372 NNKGNVSSHQQAAGNGQSGLIQRLNKN1--KPETFPALIALKORNTNLSNYSDKII 1429
Db 1574 ----EPRSNQE-----LERLANHLESEDSYTRREALAEDRE-----AKLRKVT 1614
Qy 1430 MIRPKYLVERSIGVPWBTGLDYGISQTDGTSQSSQKGFQDDFIQ-----A 1478
Db 1615 VLEEK-LVSSSNNAME-NASHQASQVQESLQEQNLNVVSKOR--DETALQLSVSQEQVKQVA 1670
Qy 1479 LGLKNTYHKGKLGSLIRIFDP-----GNELAKIKO--ASNKGEKLLKSYDLFNKYL 1530
Db 1671 LSLAN-----LQVLEHFOQEBKAMYSABELEKQKQJIAEWKKAENLEKGVISLQECLE 1724
Qy 1531 EYEKSPKIAK--GWTNIHPDQKEYPNPQKLPENLYNLVLPQWPKVTVLYNSSD----- 1582
Db 1725 EANAALDSASRLTQLDQVKEEQIEELKQNELRQEMLDDV--QKMLMSLANSSEKGVDKV 1782
Qy 1583 FITNLFV-----EPGSRD-----GSGTKLKQVIOKQVNNVADWGA--YLFVWY---- 1626
Db 1783 LMRNLFTGHFHTPKNQRHEVRLMGSILGVRRREMEQLPHD--DOGSVTRMTGWLGGGS 1840
Qy 1627 ----DKNIITNOPNVIANTADVFIKDVKELENDNTKLIAPNIT 1665
Db 1841 KSVPTPLRPNQGSVNSSSELFVKFL-ETESHPSIPPPKLS 1882

RESULT 30

US-10-335-977-8087
; Sequence 8087, Application US/10335977
; Publication No. US20040052799A1

GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 8087:

SEQUENCE CHARACTERISTICS:
LENGTH: 2384 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...2384
SEQUENCE DESCRIPTION: SEQ ID NO: 8087;
US-10-335-977-8087

Query Match 2.9%; Score 283.5; DB 15; Length 2384;

Best Local Similarity 18.5%; Pred. No. 9e-07;
Matches 394; Conservative 298; Mismatches 638; Indels 795; Gaps 105;

Qy 41 QGVISQLGID----SVAFKPSI--ANFTSDYQSVKALLNGKTFDPKSEPTDFVSKFD 94
Db 537 QG-LDQLGFNNNDHHPKEIFEPSLGTGKFTAHAPSDKNYRFITGTELDPIISANLSKFLVFNQ 595
Qy 95 FLTNNGRTVLEIPKPKYQVVIS-----EFSPEDDKE-----RFRGLFHLKEKLEDG- 139
Db 596 VIQNTALENYQFYQYDAFVGNPPYGNHKKIYSSNDKELSNESIHNYPFLGKATKELKDDGI 655
Qy 140 -----NIAQSATKFTYLLPLDMPKA----- 159
Db 656 GAFVSSWFMDAKNPKMREHIAKNAT---FLGAILRPNVPKATGAETVSDIVFPKKGVE 712
Qy 160 -----ALGOYSYIVDKNFNNL---IIHPLSNFSAQSIKPLALTRSSDFIAKLNQFN 207
Db 713 KATNQSFTKAMPYV---DKILNSLDDDTLFAQLNNRPFDFIP-----SDQLKIVNAVA 762
Qy 208 N-----QDELWVYLEKFFDLEALKANIRLOTADFSF--EKGNLVDVFPVYFIRNPNQKE 260
Db 763 NHFGPKQEKLRWYK-----DKILNSLDDDTLFAQLNNRPFDFIP-----SDQLKIVNAVA 762
Qy 261 WASDLNODOKTVRLYLRTFSPQAKTI-----LKQVYKDETFLSSI---DLKASNGTSL 312
Db 802 GKNSINLNEQTNEY----FIHPHENILGHLSLEKTRYRFEETNGEIQYKYVDLQALDESL 857
Qy 313 FANENDLKDQLDVLDDVSDYFGGSEITITNSQVXPVPASERSLKDRVKFKDQKQKPRI 372
Db 858 DLSQ-ALKQAIEKLPKDVYQY----HKTTAKTDVLIIDSSNERYQEVQKLIKNI-----L 906
Qy 373 EKPSLEYDALSFYSQLOE-----LVSKP-----NSIKDLVNAVL- 407
Db 907 ERRELVKWDLN--YFQLEQNNEMGIFLKPFTAINSKVQDSRLKAYFKIKOALNDLTSAELN 964
Qy 408 ----ARNRFLSGKNFLFDLASHLDYFVLVSKAKIKOSSITKKLFIELPIKISLKSSI 463
Db 965 PLSSDLELENKAKNLNVYDFVKKFGY-----LNENKNRK-----DIRQDL 1006
Qy 464 LGDQEPNIKTLFEKEVTFKLDNFRDVEIEKAFGLLYPGVNEELEQARKAQASFEKEKSK 523
Db 1007 YGAKVLGLEKDFEKEITPRSAKMQNIEPRQA-----QAKKAQ-IPFERTLNP 1052
Qy 524 KGLKEFSQOKEENSKAINNOEGLEED--DNTERLPENSPIOYQOENAGLGSAPDKPYM 580
Db 1053 KKELIITNAKEALIASINQKGLDLHLFIRDHFTTOSLETT----- 1092
Qy 581 IKDVQVQNR--YLLAKSQIQELIKAKDYTKLAKLLSNRHTYNIISLRLEQOLFVDNPRIPSSR 639
Db 1093 IKELLEQKLIYKDHKONGGVILANDY-----LSG-----NVKREKKE----- 1129
Qy 640 DIEKAKFVLDTKTEKNKYWOIYSSASVPFQNKWSLFGYYRYLLGLDQKOTHELVLKLGQKA 699
Db 1130 -----VKEAIN----- 1135
Qy 700 GLQFEGYENLPDSFNLLEDLKNIRIKTPLFSQKDNFKLSLLDFNNYYDGEIKAFEGFLPLF 759
Db 1136 ----QGVLEGL--EANVKOLE-----LI 1151
Qy 760 LPKELRRNSSGGSONSPWPEQEIISQFKDQNLNSQDLAQFSTKIWEKIIGDE-NEF 818
Db 1152 IPKDLKATEIMA-----NINSPW---IPTQYL-----EEFLMELSANHYEKQYGDQMTDY 1198


```

; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykend, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64245
; LENGTH: 1818
; TYPE: PR1
; ORGANISM: Mycoplasma pneumoniae
; US-10-282-122A-64245

Query Match          2.9%; Score 280.5; DB 15; Length 1818;
Best Local Similarity 19.6%; Pred. No. 9e-07;
Matches 310; Conservative 261; Mismatches 628; Indels 383; Gaps 66;

QY 56 KPSIANFTSD-----YQSVKALING-KTFDPKSSEFTDFV--SKFDFLNNGRT 102
DB 323 KASLANLTKEKERSAEXDSFERLNTALNDINMEQENALFAKLEQQOYEFERKQOES 382
QY 103 VLEIP---KYYQVVISSEPDDEKFRGLGFLHKELE-----DGNIAQSAKTFYLLPL 154
DB 383 LUKLETHKQLQKRGIEFKIESEAKSEALLIQERELLEKREIDDLTLQASLEYEQRRRT 442
QY 155 DMPKAALQGYIVDKNFNNLIHPLSNFSAQSIKPLALTRSSDFIAKLQNFNNODELWV 214
DB 443 NQ---VLKEKHQVQOQHONL-VHAKK-----LDQKHVLAEOKRI---DEROI 485
QY 215 Y-----LEKFF-----DLEALKANIRLOTADFS---PEKGNLVDPP 247
DB 486 FKLKEKATIERELEKLYLVKKQKQDKQDKENDLLIFEKQLROYQADFENEIEBKQN--ELF 543
QY 248 V-----YSFRNPQ---NOK-----EWA-----SDLN-----ODQKT 271
DB 544 ASQKSLOKSFTQKKNKEALNQAOKIAEDWAHLKONKHHADLEIFLEGEFNHLQOEKH 603
QY 272 VRLYLTETSPQAKTILDKYKDETFL---SSIDLKASNGTSLFANENDLKQDLVDL 327
DB 604 KULEARTQDNRVLSLUSARFKQKQALVKQKQSLQEQTLTAAFNKEQEAVERDKRL--ANL 562
QY 328 LDVSDYFGQSQETITNSQVFPVPSASERSLKDVRFKDQKQKPRIEKFSLEYDALSFS 387
DB 663 EKQKEMLGDKVHQFDENSINISKLAEREL--AIKFKEKELEAAQKQLSLDNNNAGLKL 720

```

```

388 QLOELVSKPNSIKDLVNATLARNLRFSLGKYNFLFDDLASHLDYFYFLVSKAKIQSSITK 447
721 QLDKLSLTERLELEASKERILDP-----YDESSRIADYSDLDQARLAEVKTL 772
448 KLFIELPIKISLSKSSILGDQEPNIKTLFEKVTFFKLDNFRD-----VEISKAFGLLYPGV 502
773 K-----NQETAAKS--ERELKVALEKLNQAKKAFQIRKQOLLEIASV 814
503 NEELEQAR---KAQRASFEK--EKSXKGLKEFQOKEENSKAINN-----EGLEEDDNIT 553
815 KQLAQKAKANLLKNOQAELDKQTBELAAFLQDOTDKLEKALKHLSVSKOELLERSFL 874
554 ERLPENSIOYQOENAGLGPDPKPYMIKDVQVQRY--YLAKSQIOELIK----- 601
875 -----LQKQREFAEHVAGFKQVHFKTQWQLSEFNKQQQSEQIKRETELKIAFAD 926
602 -AKDYTKLAKLLSNRHTYNTISLRKE-----QLFDVNPRIPISSRDIEKAEVFL 648
927 LKKDY-QLFELQKNOEQFOIEQKHKELELLAQKAELEKQKATALASQODQTVQAKL 985
649 DKTEKNKYWOIYSSASPVFQNKWSL-FGYRYLLGLDPKQTIHELKLGKAGL----- 701
986 DLARQQHELELRQNAF---NQASLSLNKOREQJTNQVULHGBELKRRHEKLTUKDRLLA 1041
702 -----QFEGVENLPSPDNF---LEDLKNIRIKTLPFSQKDNFKLSLLD 740
1042 EKEKDQHKDAEINORFKQFENYADFOAKRELOELNQIR-----RNLEQSNASLLK 1095
741 FNNYYDGEIKAPFGLPLFKELRRNSNBSGQNSNSPWEQBIISQFQD-QNLSNQ-- 797
1096 KRN-----QULTDPAALLRKVQHTQTNRVQLNTQI---KEFLLEKKNFKQASDEAA 1143
798 -----DOLAQFSTKIW---EKIIGDENPFDONNRLOYKLLKDLQSWINKTRDNLYWT 847
1144 LQKALLIKRLSRFASKLOQREALAIQKLEPKKD-----EQCKSEINNAKLQ-- 1192
848 YLGDKLKVPKKNLEAKFRQISNQLQELLTAFTYSAALSNWNNYQDSGAKSTIIFEEIAE 907
1193 ---EQFKLEKQNFDEAKQKQIEFKDQCORLDVEKRLKQKLVQLKNSKSVLYTKNRAD 1249
908 LDPKVEKVGADVYQLKPHYAIGFDDNAGKFNQEVIRSSRTIYKTSKSKLEADTDQ 967
1250 LSQOOLQHKYANLLELK-----EKLOTAKRALDKKHRAIYKQMAQFYSE 1293
968 LNQAVKAPGLQSFYLDTER-FGVFQKLAHSLAVQHKQKEKTPKPLNNDGYTIHDKL 1026
1294 LRQSKQL-LSAQKQVDDKSRLLQONQHLQNLSETKKKKQSLQSLHDK-----FQOR 1346
1027 KKPVIPOISSPEKDWFEGLNQ-NGOSQNV--NVSTFGSIESPYSTNFQEDADLQD 1083
1347 RKEAVSSILNSHK-----KLKQKEGELQGIQLKSLAKTKQIEQ-FSKLYQOREKLDQ 1399
1084 QGDSRQGNNSLDNQAGLLKQKLAILLGNQFIQYQONDKIEBEIINVEKVSLSFRV 1143
1400 RTTSLKLHRELKAQNEATAHKNREVLTEIEN---YKKELQRLATTEKSEFNNKNRLEFY 1455
1144 EFKLAKTLEDNGKTRIVLSDETMSLIWNTTIEKTPEMSAVPEVDFTKVQOYDPRTPAA 1203
1456 FRKIRNETEKKBAHIKTVLBETQK-----KRLHVEVEAKLHLQKQSIIS 1500
1204 KTKFVLKPKDOIIVDGSIGNISDKVLASIPLVHQMLRLSPVVKTIIRBLGLKTEQOQQQ 1263
1501 KQOELKEIKERSVDISH--TNKQREELNSLLHQNLKQKLAEREREINNKDSLTQKI 1558
1264 QQQQQQQQPKKA-----VRKSEE-----LETYN-----PKDEFINLPLTKAHL 1303
1559 QTAQKLEKEARILKLEKRAVEQQYQASITRLKTRNADLEKNDKNDKHLFPPLPKIN-- 1616
1304 TSLNVLNNDPNY-----KTEDLVKINEAGDHQALAFSLRANNIKLMMNPTIFA 1352
1617 -----GNDMNYPYPPYPMFYQQQKQED-----SSNQIRHLFEQOLQFM 1653
1353 DYNPPFYTNE--DWRSIDKYLNNK-GNVSSHQQAAGGQSGLIQRLNKNI-----KPE 1404

```

```
Db 1654 QOR---YENELTELRRORALLEKKLDQIQLESQLSAKNDPEKVEQMMQKLEKTEQKLS 1710
QY 1405 TFTPALIALKDRNNNTLSNSYSD 1426
Db 1711 AFDQKINALAEQINTQKAHAD 1732

RESULT 35
US-10-732-923-3306
; Sequence 3306, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Egerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCES: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3306
; LENGTH: 1818
; TYPE: PRT
; ORGANISM: Mycoplasma pneumoniae
US-10-732-923-3306

Query Match 2.9%; Score 280.5; DB 17; Length 1818;
Best Local Similarity 19.6%; Pred. No. 9e-07;
Matches 310; Conservative 261; Mismatches 628; Indels 383; Gaps 66;

QY 56 KPSIANFTSD-----YQVKKALLNG-KTFDPKSEFTDFV--SKFDFLTNNGRT 102
Db 323 KASLANLTKEKRLSAEKDSFERLNTALNDINRMEQENALFAKHLQEQYEPERKQES 382
QY 103 VLKIP---KKYQVVISPEPDEKFRGLFHLKEKLE-----DGNIAQSAKFYLLPL 154
Db 383 LLKLETHKQLKRIQEFKIESEAKSEALLIQERELLEKKREDDLLTQASLEYEQORRT 442
QY 155 DMPKAALGOYSYVDKNFNLLIHPLSNFSQAQSIKPLALTRSSDFTAKLNQFNQDELWV 214
Db 443 NQ---VLKEKHRVQOHQNL-VHAKKK-----LDQKHYLAQKRI---DEEQI 485
QY 215 Y-----LEKPF-----DLEALKANIRLOTADFS---FEKGNLVDPF 247
Db 486 FKLKEIATERRELEKLYLVKKQKQKQKQKENDLLIFEKQLRQYQADFENEIEEKON--ELF 543
QY 248 V-----YSFTRNQ---NOK-----EWA-----SPLN-----ODQKT 271
Db 544 ASOKSLQKSTQKKNKEAELNQAKQIAEDWAHLKQKHHADLEIFLGEFNLQOEKH 603
QY 272 VRLYLRTFSPQARTILKDYKDETFPL-----SSIDLKASNGTSLPANENDLKQDLDVL 327
Db 604 KLEARTQFNVRVSLLSARFKQAELVKQKQSLQEQITAAFNKEQEAVERDWKDL--ANL 662
QY 328 LQVSDYFGGQSEITNSQVQKVPASERSLKRVKPKQKQKQKPRIEKFSLYEYDALSFYS 387
Db 663 EKQKEMLGDKRVHGFQDENSINLSKLAEREL--AIKFEKELEAAQKQLSLDNNNAGLKL 720
QY 388 QLOELVSKPNSIKDLVNATLARNLRFSLGKYNFLFDDLASHLDYFLVSKAKIKQSSITK 447
Db 721 QLDKLSUKTERLELEASKERILDF-----YDESSRRIADYSDDIQAARLAEVKTLE 772
QY 448 KLPFLPIKISLSSILGQEPNIKITLFEKEVTFKLDNFRD-----VEIEKAFGLLYPGV 502
Db 773 K-----NQETAAS--ERELKVALEKLNQAKKAFQIRKQQLLEIASV 814
QY 503 NEELEQAR---KAQASFEK--BKSKKGLKFPQKQKSEKSKAINN-----QEGLEEDNDIT 553
Db 815 KOQLAQKANLLKQQAELDKQTELEAAFLQEQDTDKKELEKALHVSQKQELLERSFL 874
QY 554 ERLPENSIQYQENAGLGASDPKPYMIKDQVQORY-YLAKSQIQELIK-----601
```

RESULT 36
US-10-335-977-8088
; Sequence 8088, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al

```
Db 875 -----LQKQREFAEHVAGFKRQVHFKTQMQRULSEFNKQQQSEQIKRETELKTAFAAD 926
QY 602 -AKDYTKLAKLSNRHTYINISLRKE-----QLFDVNPRIIPSSRDIEKAKFVL 648
Db 927 LKQDY-QLFELQKNQEFQOIQQKHKELELLAQKQAELEKQAEKATALASQODTVOAKL 985
QY 649 DKTERKNYQWIIYSSASPVFQNKWSL-FGYRYVLLGLDPKQTTHELKVKQKAGL-----701
Db 986 DLARQOHELELRQNAF---NQASLSLNKQREQLTNQVVLHGLKKRHEKLTLDKRLLA 1041
QY 702 -----QFEGYENLPDPFN-----LEDKNIIRIKTFLFSQKQNFKLSLLD 740
Db 1042 EKEKQHKDAEINQRFQFENEYADFOAKKRELQELNQIR-----RNLEQSNASLLK 1095
QY 741 FNNYDGEITKAPEGLPLFLPKELRRNSSGGSQNSNSPWEQEIISQPKD-QNLSNQ--797
Db 1096 KRN-----QLTLDPALLRKVQHTQTNRVQLNTQI---KEFLEKKNFQKASDEAA 1143
QY 798 -----DQLAQFSTKIW---EKIIGDENEFQNNRLOQLKLDQESWINKTRDNLWYT 847
Db 1144 LQKALLIKELRSFASKLQKREALAIQKLEFKRD-----EQKSEINNNAKLQ---1192
QY 848 YLGDKLKVKPKNLEAKFRQISNLQELLTAFYTSAAALSNNVYQDSGAKSTIPEIEAE 907
Db 1193 ---EQFKLEKQNPDEAKQKQLEFKDQCQORLDVEKRLKQKLVOLKKNLSKSYLTYNRAD 1249
QY 908 LDPKVKKEKVGADVYQLKPHYAIGFDDNAGKFNQEVIRSSRTIYLYKTSQSKLEADTIDQ 967
Db 1250 LSOQQLQHKYANLLELK-----EKLQAKRALDKKHRAIYGMAQFVSE 1293
QY 968 LMQAVKNAPLQOSFYLDTER-FGVFQKATSLAVQHKQKEKTLPKKLNNDGYTLHDKL 1026
Db 1294 LRQEKQL--LSAQKQVDDKSRLLQNRHLQNSLSETKKKQSSLEHDINK-----FDOR 1346
QY 1027 KKPVPQIPISSSEKDFEGLNQ-NGOSQNV-NVSTFGSIIESPYFSNFOEDADLDD 1083
Db 1347 RKEAVSSILNSHK-----KLQKQEGELGIIQKLSLKTQIEQF-FSKLYQOREKLDRO 1399
QY 1084 QGDDSRQGNNSLDNQEAGLLKQKLAILLGNQFIQYQQNDKEIFEIINVEKVSLSFRV 1143
Db 1400 RTTSLKHLRELKAQNEATAHKNREVLEIEN---YVKELQRLTTSEKSEFDDNNKRLFEY 1455
QY 1144 EFKLAKTLEDNKTIRVLSDETMSLIVNTTIEKTPMSAVPEVFDTKWVEQDPRTPLAA 1203
Db 1456 FRKIRNEIEKGAHKTVLEETQK-----KRHLVTEAVKHLQKQSIIS 1500
QY 1204 KTKFVLKPKDQIPVDGSGNISDKWLASIPLVIHQOMLRSLSPVVKTIRELGLKTEQOQOQ 1263
Db 1501 KQELKEIKERYSRDISH--TNKQREELNSLHQNKLQKNLAEREREINNKKDSLTKI 1558
QY 1264 QOQOQOQPOKKA-----VRKEBE-----LETYN-----PKDEFNILNPLTKAHL 1303
Db 1559 QTAQKQKSEKARILKLEKRAVEQYQAHITRLKTRNADLEKNDKHLFPPLPKIN--1616
QY 1304 TSLNVLNNDPNY-----KTEDLVKNEAGDHQALAFSLRANNIKRLMNTPTIFA 1352
Db 1617 -----GNDMNPYPYPMFPYPOKQED-----SSNQIRHLFEQOQLQM 1653
QY 1353 DYNPPFYNE--DWRSIDKYLANK-GNVSSHQOAGGQSGLIQRLKNKI-----KPE 1404
Db 1654 QOR---YENELTELRRORALLEKKLDQIQLESQLSAKNDPEKVEQMMQKLEKTEQKLS 1710
QY 1405 TFTPALIALKDRNNNTLSNSYSD 1426
Db 1711 AFDQKINALAEQINTQKAHAD 1732
```

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: Windows NT 4.0

SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977

FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002

FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragoras, Amy E.

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 8088:

SEQUENCE CHARACTERISTICS:

LENGTH: 2440 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...2440

SEQUENCE DESCRIPTION: SEQ ID NO: 8088:

US-10-335-977-8088

Query Match 2.9%; Score 280.5; DB 15; Length 2440;

Best Local Similarity 18.5%; Pred. No. 1.4e-06;

Matches 395; Conservative 316; Mismatch 678; Indels 697; Gaps 100;

QY 41 QGVISQLGID---SVAFKPSI--ANFTSDYQSVKALLNGKTFDPKSSSEFTDFVSKFD 94
DB 593 QG-LDQLGFNDNHPKKIPEPSLGTGKFTAHAPSDKNYRFICTELDPISANLSKFLYPNQ 651
QY 95 FLTNGRTVLEIPKYYOVVIS-----BFSPEDDKE-----RFLGFLHKEKLEDG- 139
DB 652 VIQNTALENYQFYQYDAFVGPNPYGNHKKIYNSDNKELSNESIHNYFLGKAIKELKDDGI 711
QY 140 -----NIAQSAKTFIYLLPLDMPKA----- 159
DB 712 GAPVSSWFMDAKNPKMREHIAKNAT---FLGAIRLPNSVFKATGAETVSDIVFFKKGVE 768
QY 160 -----ALGOYSYIVDKFNFL---LIHPLSNFSAOSIKPLALTRSSDPIAKLNQFN 207
DB 769 KATNOSFTKAMPYY-----DKILNSLDDDTLFLALQNNRFDSPF-----SDQLKIVNAVA 818
QY 208 N-----QDELWVYLEKFPDEALKANIRLOTADFSF--EKGNIIVDPFVYSFRNPNQKE 260
DB 819 NHGFGQKELQRYEK-----IDTANFGYSTQDYKIIKDFI-----DKV 857
QY 261 WASDLNQDQKTVRLYLRTFSPQAKTI-----LKQYKDETFLLSI-----DLKASNGTSL 312
DB 858 GKNSININEQTLNEY----FIHHPENILGHLSLEKTRYRFETNGEQIYKYDQLQALEDESL 913
QY 313 FANENDLKQDLVDLLDVSDFYGGQSETTTSNSQVFPVAPASERSLKDVRKFKKQDQKPRI 372

DB 914 DLSQ-ALKQAEIKLPKDVQY-----HKYTLTKDVLIDSSNERYQEVOKLKN-----L 962
QY 373 EKFSLEYDALSYSQLOE-----LVSKP-----NSIKDLVATL- 407
DB 963 ERRELKVDNL--YFQLEQNNMGIFLKPTKINSKVQDSRLKAYFKIKDALNDLTSAELN 1020
QY 408 ----ARNLRFSLGKYNFLPDDLASHLDYFLVSKAKIKOSSITKCLKFIEPIKISLKSSI 463
DB 1021 PLSSDLELENKRAKLNLVYDFVKFGY-----LNENKRX-----DIRQDL 1062
QY 464 LGDQEPNIKTLFBEKVTFFKLDNFRDVEIEKAFGLLYPGVNEEIEQAKAQAQASFEKSKS 523
DB 1063 YGAKVLGLEKDFEKEITPRSAKQNIETPROA-----QAKKAO-IFPERTLNP 1108
QY 524 KGLKBFSSQKKEENSKAINNOEGLE-----EDDNITERLPENSPIQOENAGL 571
DB 1109 KXELIITNAKEALIASINOKGGLDHFIRDHFTTQSLTETIKELLEQKLYKHOKHGGY 1168
QY 572 GASPD-----KPYMIKDVQQRVYLAKSOIQ--ELIKAKDVTKLAKLLSNRHTYNIS 621
DB 1169 ILANDYLSGNVVKLKEVKRAINQGVGLEANVKOLELIIIPKD----- 1211
QY 622 LRLKEQLFDVN-PRIPSSRDIEKAFVLDKTEKNKYQIYSSASPVFNKWSLEGGYRYL 680
DB 1212 LKATEIMANINSPWIPT---QVLEBFLMELSANHYEKQYGDKMTDYQ----- 1255
QY 681 LGLDPKOTIHELKVLGQKAGLQEGEYENLPSDFNLEDLKNIRKTIPLFSOKDNFKLSLLD 740
DB 1256 ----LSNLKEDIKIEHLSG---AYEVFVRNNELNELYGIRHKDKPHSKVPESLUNK 1306
QY 741 FNNYVDGEIKAPFGLPLFLPKELRRNSGSGNSNSPWEQIIISQFQDNLSNQ--D 798
DB 1307 VLNNKLSVKYQV-----DPNDPKKEIFIIDEQSNLARQKAE 1345
QY 799 QLAQ-FSTKIWEKIIGDENEFQNNRLQYKLLKDLQESWINKTNDLYWTVLGYDKLVKVP 857
DB 1346 ELKEAFKDWIVK-----DYSRRTHLE-QIYNDTFNNFLKTYD-----GSQLELEG 1390
QY 858 KN-----NLEAKFRQISN---LOELLTAFYTSAAALSNNWNYQDSGAKSIIIFEEI 905
DB 1391 FNYHISLRPHQKNAIFRITQDRAVCLDHQVAGAGTLCALIASCMBEQKRMGLVKNKTLI--- 1446
QY 906 AELDPKVKKGVADYQYQLKFHYAIGFDDNAGKFNQEVIRSSRTIYLTSGSKSLEADTI 965
DB 1447 -AVPNHLTKQMGDFYKAYPNANVLVVDSDKDTTKE-----RELLFNQIANNYDAVVI 1499
QY 966 DQLNQAQVKNAPLGLQSFYLDTERFGVFQKLA TSAVQHOKKEKTLPPKLNNDGYTLIHDK 1025
DB 1500 AHTHLELLSNPRGI-----IEELKEEELVNAEKNFERQELAYKN 1538
QY 1026 LKKPVIPOISSPKDWPFGKLNQNGQSONVNVSTFGSIIIESPYSTNFQEDADLDQDQO 1085
DB 1539 ----NPRETKKPNERAFKNKLDK-----IRAKY-----DAILEKQGS 1571
QY 1086 --DPSRCQNSLNDQEAELK-----OKLAILLGNQ-----FI--QYQQNDK 1124
DB 1572 HIDISQMGIDNLIIVDEAHLFNALAFETSMEXIA-GLGNQOQSNRARDLFIKTRYLHQNDK 1630
QY 1125 EIEF-----EIIINVEKVSLSFRVFKLAKTLEDNGKTIRVLSDETMSLIIVNTTIKTPM 1180
DB 1631 KIMFLTGTPIAN--SLSEMYHLQRYLTPDVLKERGL----- 1664
QY 1181 SAVBEVFTKWEQY-----DPRTPLAKT-KFVLKFKDQIPVDGSGNISDKWLASIPLV 1234
DB 1665 ----BFFD-DWAKTYGEVNDFELDTSAQSYKVMNRFKFSVDQO---LSTMYRAFADIV 1716
QY 1235 IHQQMLRLSP-----VWKIRE-----LGLKTEQOQOQOQO---QOQOQPOK 1274
DB 1717 SNDDLKKNHPFVKYGDKINVVVKSEBEVAQFIGVALENGKYNESIIDRMQKCEGK 1776
QY 1275 AVRKEELETYNPKDQFNILNPLTKAHLTLIS-NLVNNDPNYKIBDLKVIKNEAGDHOLA 1333

Db 1777 KSQKQD-----NLSCTTDARKVALDYRLI--DPNAKVEK-----BFSKS 1815
Qy 1334 FSLRANNIKRLMTPITFADYNPFYFYNEDWRSDIKYLNKNVNSVSHQQAAGNQGSL 1393
Db 1816 YAMAKNIYENLETHATKGTQLGFI-----CLSTPKTHSQKVS----- 1853
Qy 1394 IQRLNKNIKPDETTTPALIALKDRNN---TWLSVNSDKIMIKPKYLVERSIGVPMWSTGLD 1450
Db 1854 LEALDNAHETENKNPL-----DKAQELLESLSYDEKGNLIAP----- 1891
Qy 1451 GYIGSEQTKDGTSSSQKQDFDODFIOALGKNTYHGKLGSLRIFDPGNEIAKIKDAS 1510
Db 1892 -----SKKLESELKEKAESVNLDEEIAKGCSDVYSVLRHLVQMGPONEIAFIHDA- 1946
Qy 1511 NKKEEELKLSYDLFKNLYNEYKK-----SP-KIAKGWTNI-----HPDQ- 1550
Db 1947 --KTEE--QKQDLFKK-LNRGGVRVLGSPAKMGVG-TNVQERLVAMHEDLDCPWRDEL 1999
Qy 1551 -----KEYPNQKLPENYLNVL-----NQPW-----KVT 1576
Db 2000 LOMEGRGIRQGNILHQNDPENFRMKIYRYATEKTYDSRMWQIITETSKGIEQPRNAHKL 2059
Qy 1577 LYNSSDPITNLVFEPEGSDRSGTK-----LQOV-IQOVNNYADMGSAYLTFWY 1626
Db 2060 LNELEDF--NN-----GSSNASEKABATGNPLIIEVKURAEIKSEESKIYKAFNKEHYF 2112
Qy 1627 DKNIITNPQNVITANIADVPKIVKELED--NTKLIAPNITQ--WPNNISGSKKPY----- 1679
Db 2113 NEESLKN-----NASKLDYLKQELKOLETLQRSVIIPTHTTEIKLYDLKNEESKDYELIKV 2167
Qy 1680 ---KPTVFFGNWENENSNMSQATPTWEKIRE-----GFAOLAKSSFDQKTRTF 1727
Db 2168 KEVEPL-----KENASMEEL---THKKLKEQNKOIABQNKELDAIKQFASNLNTL 2217
Qy 1728 VLTNAPLPLWKYGPLGF-----QNGPNFKTDWRLVFOND----- 1763
Db 2218 FVNEEDYKLEY--KGFVNVNAKYQVFEFSLSPKDNPNIAVSPSNVYKNDTINWFSS 2275
Qy 1764 -----DNOIAALRVQDRPEKSSE 1783
Db 2276 YNFCABIKPGLKRLDNAITKL-----PEKIKE 2304

RESULT 37

US-10-732-923-3353
; Sequence 3353, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3353
; LENGTH: 1478
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1478)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-3353

Query Match 2.98; Score 279.5; DB 17; Length 1478;

Best Local Similarity 18.3%; Pred. No. 7.6e-07;
Matches 313; Conservative 268; Mismatches 607; Indels 519; Gaps 76;

Qy 353 SERSLKDRVFKKDOQKPRTEKFSVLYSDALSFSYQELVSKPNSIKDLVNAATLARNLR 412
Db 21 SRKRESKXVENADD-----ENLAIKERELINQLNELNIISEKEMHSDAI-----KNMK 70

Qy 413 FSLGKYNFLDLDLASHLDYVFLVSKAKIKQDSITKFLFIELPIKISLSSILGDQBPNI- 471
Db 71 HQLDEFQ-----DQNNYI 84
Qy 472 --KTLPFEKEVTKLDNFRDVEIEKAPGLLYPGVNEBELEQARKAQRASFEKESKKGJKEP 529
Db 85 NELRILEKLDLTKRD-----EIEK-----FNKLEBINRREESLAYKE-----NEF 125
Qy 530 SQQEENSKAINNOEG--LEEDDNITERLPENSPIQ---YQOENAGLGASPKDPYMIKDV 584
Db 126 IQIKEE-----INNRERDIIDREHEITEKVNELSYKEKELVEKEHELLNKESDILKNNEF 181
Qy 585 QNQRYYLAKS-----QIQEL-----TKADYTKL-----AKLSN 614
Db 182 INNNEELARKBEELNTEQIQIBELKNDLILKPEHLHNEQDIKIKE-TKLKTIYKKLKN 240
Qy 615 RHTYNISLRLKEOLF-----DVNPRIPSSRDI-----EKAKFVLDKTE 652
Db 241 NINDLLDVSNDENLFSRTRSNINGNENENSNISMRKIAELDNLENQSENLNLDLSLE 300
Qy 653 KKKYWOIY-----SSASPVFNKWSLFGYYRYLLGLDP-----KOTIHELVLK 695
Db 301 YTYTFNEFKKIKNLKQDLQSRKELEFFETLEIEKREKEEYRNELDNKLLHAEQLN 360
Qy 696 GQKAGLQFEGYENLPDFNLEDLKNIRIKTLPFSQKNFKLSLLDPNNYVDGEIKAPEFG 755
Db 361 KKKLDDQLSRYKN--DDEHI--IKSLKESEEIINEKNTLILEL-----QQKLAQASYE 409
Qy 756 LPLFLPEKLRRSNSGSGSNSPWEQEIISOPKDNLSNQDQLAQFSTKIWEKIIGDE 815
Db 410 ISMIENKSNKNNNSYNGER-NSEYEKKL-----EELNN-----ITNSYEKINEL 455
Qy 816 NBFQDNRLQY--KLLKOLQESWINKRDNLYWYLGDK---LKVKPKN-----L 861
Db 456 NKEKENITQPFMEKIMDD--EEGINKLKEE---LNDANALIVLNKNKNELNYYNIEM 508
Qy 862 EAKFROISNLQELLTAFYTSAAALSNWNNYVQSGAKSTIIFEEIA----- 906
Db 509 EQANKDMRDDIDILLA--NIDKLNDEKNTVEKEQELKYNELKINYEKHKCNKFPN 566
Qy 907 ELDPKVKVKGADVOLKEHYAIGFDDNAGKNOEVISSRTIYLTSGCKSLEADTID 966
Db 567 MLPPKMKKI---EYKGRNSKEGINNSQNSVSIKDHESL-----GNSSEETCDID 619
Qy 967 QLNQAVKNAAPLGLQSPY---LDTERFGVFKLATS LAVQHKKQKTLPKKLANDGYTLI 1022
Db 620 KIVNVVNGQNGKRASIKINNILNEKDMVIBSLNEKLLYVEKDNAGN--HEIINDYKSKI 677
Qy 1023 HDKLLKPVIPQISSPEKDWFEKG-----LNQOQSONV-----NVSTFGSIIE 1066
Db 678 DELLSKLN-----FEGNMPFFIKINEIINNENNYTHQYVJVLNLLNAILYRKIIT 736
Qy 1067 SPYFSTNFOEDADLDQGDSDRSQGNNSLDNQA---GLLKOKLAILLGNQPT----- 1116
Db 727 KIFIENQFYKEIIQIKYEIFDIISKNTSNKKKETLKLMSIKKSGSPKGGHSHSNYPL 786
Qy 1117 -----QYYQONDKEIEFEIINVEKVSBELSPRVEPKLAKTLEDNGKTIIRVLSDETMSL- 1168
Db 787 HMSDAIHKYONPXDBEQSGMDLIEPTNW-----IIKAVNDEPDTPENICNELYNLN 839
Qy 1169 -----IIVNTTIKTPENSAPVEPFDTKWVQYDPRTPLAOK 1204
Db 840 KSEEMDIMYKLVNLEYSYNSIINIINIKTKNISNI-----EKSELKKN 886
Qy 1205 TKFVLKFKDOIIPVDGSGNISDKWLASIPLVHQMLRL---SPVVKTI-----RE 1251
Db 887 MKLLKKKXNSLNDFLQNVEDVDKLLILVKSEQNEELLVTENEELKNLYKELNDEYNEK 946
Qy 1252 LGLKTEQEQOQOQOQOQOQOQOQKAVRKEBELETYPNPKDEFNLTPLTKAHLRLTSLNVN 1311
Db 947 LNLKQNEYOIKNLQBLIEKNEKNTKTEINEF-LKTDLDYLTLSLEQANQSLTNLSE 1005

```

QY 1312 DPNTKI-----EDLVKIKNAGDHOAFSLRANNIKELMNTPTITPADDYNNFFYYN-EDW 1364
D 1006 NEKNKIALKOLTEENIVLKNQIEDKEENIEYLTOKIKNSDOVISSELKEFEMLIKVKETY 1065
QY 1365 RSIDKYLNNKNSVSHOQAAGGNGGLI--QRLNKNIKPTEPTPALIALKARNTNLS 1422
D 1066 DSIELNRKGN--NHTKGHIENSTTLYDHNFYEDIDQVELLGVISKLENN-- 1119
QY 1423 NYSKIMIKPKYLVERSIGVPMSTGLDGYTGE-----QTKGTSSSSOQKQFDQ 1473
D 1120 NLKEEDMLKN-----DFYILSEKNHELEBIIQKNDLPKIKYNDELIDKN 1163
QY 1474 DFIQALGLKNTYHCKLGLSIRIPDPGNELAKIKDASNNKGEELKLSYD----- 1523
D 1164 DIIKENLINERNKYKI-KCIQIID-----ICFNKDFSIIDIREKIVAFENDDSEIINII 1218
QY 1524 -----LFXNYLNE-----YEKKS-----PKIAGWTNI--HPDQKEYPNPNKLPENYL 1565
D 1219 NCHRSMLNNTNEEFGYKKNPNKPNARHSRIIRGTSKVCVADNNQY-NENESMKORLS 1277
QY 1566 N-----LVL--NOPWKVTLX-----NSSDIFITNLV 1589
D 1278 NORKDDAAKTHQENITHEEMSKONEELIKKNEQIWLKNKYIEDLNKEITNKDYII-MKS 1336
QY 1590 BPEGSDRSGTKLVKIOVKQVNNVADNGSAYLTFWYDK-NIITNOPNVITANIADVPIK 1648
D 1337 QOESADK-----NDVLEKNQYIIFLKDQOIKLLCN--NIDENNLFDI--- 1376
QY 1649 DVKELEDNKTLPAPNITQWPNPISGSKFKYPTVFFGNWENENSSMNSQAQTPTWKIR 1708
D 1377 -----DNVNL-SPTLR-----SPFKRSYNNRCSFVEVDEKSNHGHE-----N 1415
QY 1709 EGFA-LQALKSSFDQ--KTRTFVLTTNAPLPLMKYGLPLGFQNGPNFKTQDWRLVFQNDN 1765
D 1416 ENFASRQRRRTFDDQSRKRLSEFTS-----QEQYNF-----YDEN 1451
QY 1766 QIAALRVOEQDRPEKSSDKOKKWK 1792
D 1452 EV---KIBEMNLKMDLNDKLKQKRNK 1475

RESULT 38
US-10-661-809-23
; Sequence 23, Application US/10661809
; Publication No. US2004010191A1
; GENERAL INFORMATION:
; APPLICANT: HOOK, Magnus
; TITLE OF INVENTION: BIOINFORMATIC METHOD FOR IDENTIFYING SURFACE-ANCHORED PROTEINS FR
; FILE REFERENCE: P07741US01/BAS
; CURRENT APPLICATION NUMBER: US/10/661,809
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: 60/410303
; PRIOR FILING DATE: 2002-09-13
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 10203
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-661-809-23

Query Match 2.8%; Score 275.5; DB 16; Length 10203;
Best Local Similarity 18.5%; Pred. No. 2.1e-05;
Matches 375; Conservative 319; Mismatches 758; Indels 573; Gaps 93;

QY 74 LNKGT-----FDPKSEFTDFVSKFDFLTNNGRVLETPKKYQVVVISFSPEDDKERPR 127
D 3189 LNMGTQESINNYTTKEREAGNIASSADTIINNGASIE-----QITEN----- 3231
QY 128 LGFHLEKLEKEDGNIAQSATKFIYLLPLDMPKALGOYSYIVDKN-----ENNLIITH 178
D 3232 -----KIRVEEATNALNEAKHQLTADTTSLKTEVRLRRGDTNNKPSVSAAYNN-TIH 3285

```

```

QY 179 PLSNFAQSISKIPALTRSSDFTAKLNQFNNDLWVYLEKFFDLKALKANIRLOTADFSF 238
D 3286 SLOSETOT-ENPANTIINKPISVEVNNALH-----EVNQLNQLT----- 3327
QY 239 EKNGLVDPFYYSFIRPNQKQWASDL-NQOQKTVLYLRTSFSPQAKTILKDYKYKDET 297
D 3328 DTINLLQP-----LANKESLKEARNLESKINETVQTDGWTQOS-----VENYK----- 3371
QY 298 FLSSIDLKASNGTSL--FANENDLKDQ-----LDVLLDVSDVFGQSETITSNSQ 346
D 3372 ---QAKIKAGNESSIAQTLINNGDASDQEVSTIEKLNQKLSLTN-----SINHILTWNKE 3424
QY 347 VKPVPASERSLKORVKPKDQKPRIEKFSLYSDALSFYSQQLQELVSKPNSKNDLYNAT 406
D 3425 --PLETAKNLOQANI-----DQKPSD--CMTQOSVQSYERKLEQAKDKINSI---NNV 3471
QY 407 LARNRFSLGKYNFLFDDLASHLDYVFLVSKAKIKOSSITKKLFIPLPIKLSKSSILGD 466
D 3472 LANNPVNAIRTN-----KVETEQIINNEL-----TOAQKGLTVD 3505
QY 467 QEP--NIKTLPFK-----EVTFDLDNFRDVEIEKAFGLLYPGVNEELEQAR-K 511
D 3506 KQPLINAKTALQOSLDNQPSPTTGTGTEATIQYNAKROKAEQ-----VIONANKIITENAQPS 3561
QY 512 AQRASPEKSKKGLKEFSQ-----QKEENSKAINNQEGLEBDDNITERLPENSPI-- 562
D 3562 VQOVSDEKSKVEQALSELNNAKSALRADKQELQOAYNQ---LIQPTDLNNKPKASITAYN 3618
QY 563 -QYOENAGLGAS-----PDKPYMIKDVQNYRYLAKSQIOELIKAKDYTKLAKLLS 613
D 3619 QRYOQFSNELNSTKNTDRIILKEQNPVSADVNN-----ALNKVREV--QOKLNEARALLQ 3671
QY 614 NRHTYINSLRKEOLFVDNPRIPSS-----RDIEKAKFVLDKTEKN 654
D 3672 NKEDNSALVRAKEOLOQAVDQVPSTEGMTQOTDDYNSKQAAQOESKAAQVINDGAT 3731
QY 655 KYWQIYSSASPVFQNKWSLFGYRYLLGLDIPKQTIHBLVKLGKAGLQFEGYENLPSDFN 714
D 3732 T--QOISNAKTNVERALEALNNAKTGLRADKEELQNAYNQLTQ-----NIDTSGKTPASIR 3785
QY 715 LEDLKNIRIKTLPFSQKDNFKLSLLDFNNYD-----GEIKA--PEFGLPLFPLPELRR 766
D 3786 KYNEAKSRIQTQIDSAKNEANSILTNNDNPQVSQVTAALNKIKAVQPELDKAIAMLNKEN 3845
QY 767 NSSNSGGSQ-----NSNSPWE--QELISQKQNLNQDQLAQFSTKIWEKII--GDE 815
D 3846 NNALVQAKQQLQOQIVNEVDPTQGMTTDTAMNYKSKKREADEIQKA-----QOIIINNGDA 3900
QY 816 NE---FDQNNRLOYKLLKOLQESMINKTRDMLYWTYLGDKLVKPK-----NNLEAKPRQ 867
D 3901 TEQOITNETNRVNOAI-----NAINKAKNDL-----RADKSQLENAYNQLQNVDTNGKK 3950
QY 868 ISNLQELLTAFYTSAAALSNMNYQDSKAKETIIFEEIAELDPKYKEKVG-----DVYQL 923
D 3951 PASIQOYQAA--RQAITQYN-----NAKSE-AHQIILENSNPSVNEVAQALQKVEAVQL 4001
QY 924 KFHYAIGFDDNAGKPNQEVIRSSRTIYLTGSGSKLEADTIDQLN-----QAVK 973
D 4002 KVNDIAIHILQN--KENNSALVTAKNLOQSVNDQPLTTGTMQDSINNYEAKRNEAQSAR 4059
QY 974 NAPLGLQ-----SPYLDTERFVGFKLATSFLAVQHKKQKKTLPKKLNDGVTYLHDKLKK 1028
D 4060 NAEAVINNGDATAKQISDEKSKVQALA-----HLNDAKQOLTADITTEL 4103
QY 1029 PVIPQISSSPKQWPEGKLNQNGOSQNV---NVSTFGSIIESPYPSTNFQEDADLDQDG 1085
D 4104 QTAVO-----QLNRRGDTNNKPPRSINAYNKAIQSL----- 4134
QY 1086 DDSRGNNSLNQEAGLLKQKLAILLGNQFTQYQOONDKEIEFEIINVEKYSE-----L 1139
D 4135 --ETQITSKONANAVIQKPIRTVQEVNNALQOQVNNALQOQLTEAINQPLSNNDALKAA 4192

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2005, 20:50:32 ; Search time 217 Seconds

(without alignments)
3348.953 Million cell updates/sec

Title: US-10-607-631-8

Perfect score: 9732
Sequence: 1 MNKNKSTLLATAAAIIGST.....TNNAFNNVKEFNISKIVE 1879

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9598	98.6	1879	8 ADI28615	ADI28615 Mycoplasma
2	9435.5	97.0	1878	8 ADI28627	ADI28627 Mycoplasma
3	475	4.9	1032	8 ADI28611	ADI28611 Mycoplasma
4	415	4.3	1001	8 ADI28617	ADI28617 Mycoplasma
5	325	3.3	4688	6 ABU48941	ABU48941 Protein e
6	305.5	3.1	1875	6 ABR53560	ABR53560 Protein s
7	305.5	3.1	1875	7 ADK64380	ADK64380 Disease t
8	305.5	3.1	1875	8 ADS43855	ADS43855 Bacterial
9	303	3.1	2633	4 ABG06505	ABG06505 Novel hum
10	302	3.1	2663	4 AAM39097	AAM39097 Human pol
11	302	3.1	2688	4 AAM40883	AAM40883 Human pol
12	301	3.1	2663	4 ADQ17932	ADQ17932 Human eof
13	299	3.1	2017	4 ABG06301	ABG06301 Novel hum
14	295	3.0	1639	2 AAW54145	AAW54145 P. falcip
15	295	3.0	1639	5 AAQ29345	AAQ29345 Plasmodiu
16	295	3.0	1654	1 AAP50777	AAP50777 Sequence
17	291.5	3.0	896	2 AAR05591	AAR05591 The 36 kD
18	291	3.0	5005	6 ABU48947	ABU48947 Protein e
19	290.5	3.0	1679	8 ADS43650	ADS43650 Bacterial
20	290.5	3.0	1979	3 AAB18171	AAB18171 Plasmodiu
21	288	3.0	1948	5 AABP73774	AABP73774 Candida a
22	286.5	2.9	1979	7 ADB75595	ADB75595 Protease
23	281	2.9	1805	6 ABU35589	ABU35589 Protein e
24	281	2.9	1805	7 ABO23572	ABO23572 Mycoplasma
25	281	2.9	2441	3 AAB18161	AAB18161 Plasmodiu

26	280.5	2.9	1818	6 ABU36321	ABU36321 Protein e
27	280.5	2.9	2440	2 AAW20828	AAW20828 H. pylori
28	276	2.8	3259	7 ADE56037	ADE56037 Human pro
29	276	2.8	3259	7 ADE56033	ADE56033 Human pro
30	275.5	2.8	1663	2 AAR46608	AAR46608 Plasmodiu
31	275.5	2.8	10182	5 ABP38314	ABP38314 Staphyloc
32	275.5	2.8	10182	8 ADO84851	ADO84851 S. epiderm
33	275.5	2.8	10203	8 ADS04803	ADS04803 Staphyloc
34	274.5	2.8	1997	3 AAB18287	AAB18287 Plasmodiu
35	273.5	2.8	1690	4 ABB61144	ABB61144 Drosophil
36	273.5	2.8	1690	4 ABB61173	ABB61173 Drosophil
37	272	2.8	1516	3 AAB18195	AAB18195 Plasmodiu
38	271	2.8	3225	7 ADJ68448	ADJ68448 Human hea
39	270	2.8	6641	6 ABU42656	ABU42656 Protein e
40	269.5	2.8	1464	8 ADP25450	ADP25450 Plasmodiu
41	262.5	2.7	2295	3 AAB18180	AAB18180 Plasmodiu
42	262.5	2.7	2681	6 ABU19025	ABU19025 Pathogen
43	262.5	2.7	5024	4 AAG82935	AAG82935 S. epider
44	262	2.7	3418	7 ADJ68372	ADJ68372 Human hea
45	261.5	2.7	5171	7 ADJ70881	ADJ70881 Human hea

ALIGNMENTS

RESULT 1

ADI28615
ID ADI28615 standard; protein; 1879 AA.

XX AC ADI28615;

XX XX 22-APR-2004 (first entry)

XX DE Mycoplasma hyopneumoniae immunogen protein C28-MHP545.

XX KW Pneumonia; vaccine; diagnosis; C28-MHP545; immunogen; immunostimulant;
XX KW antiinflammatory.

XX OS Mycoplasma hyopneumoniae.

XX XX WO2004003161-A2.

XX XX 08-JAN-2004.

XX XX 27-JUN-2003; 2003WO-US020460.

XX XX 28-JUN-2002; 2002US-0392632P.

XX XX (IOWA) UNIV IOWA STATE RES FOUND INC.
(NSWA-) NSW AGRIC.

XX PI Minion CF, Mahairas GG, Djordjevic SP;

XX DR WPI; 2004-083044/08.

XX DR N-PSDB; ADI28614.

XX PT New immunogenic Mycoplasma hyopneumoniae polypeptide, useful in eliciting
an immune response and in treating or preventing enzootic pneumonia.

XX PS Claim 5; SEQ ID NO 8; 81pp; English.

XX CC The present sequence is the protein sequence of C28-MPH545, an
immunogenic polypeptide from Mycoplasma hyopneumoniae strain 232. The
invention provides M. hyopneumoniae polypeptides and nucleic acids
ADI28608-ADI28627, including C28-MPH545, as well as vectors and host
cells. Compositions containing M. hyopneumoniae polypeptides and nucleic
acids are provided for use in methods of treating pigs to prevent
enzootic pneumonia. Diagnostic tests for detecting M. hyopneumoniae
infection in swine herds are also provided.

XX SQ Sequence 1879 AA;

Query Match 98.6%; Score 9598; DB 8; Length 1879;

Best Local Similarity 98.8%; Pred. No. 0; Matches 1857; Conservative 8; Mismatches 14; Indels 0; Gaps 0;		
Qy	1	MKNKSTLLATAAAIIGSTVFGTVGLASKVKYRGVNPQTQVISOGLGIDSVAFPSTA 60
Db	1	MKNKSTLLATAAAIIGSTVFGTVGLASKVKYRGVNPQTQVISOGLGIDSVAFPSTA 60
Qy	61	NFTSDYQSVKALLNGKTFDPKSESEFTDFVSKPDFLTNNGRVTLEIPKKYQVVISFSP 120
Db	61	NFTSDYQSVKALLNGKTFDPKSESEFTDFVSKPDFLTNNGRVTLEIPKKYQVVISFSP 120
Qy	121	DKERFLRGHLKEKLEDGNIASATKFIYLLPLDMPKALGOYSYVDKFNFNLIHPL 180
Db	121	DKERFLRGHLKEKLEDGNIASATKFIYLLPLDHPKALGOYSYVDKFNFNLIHPL 180
Qy	181	SFNSAOSIKPLALTRSDFTAKLNQFNQDELAVYLEKFPDLBALKANRLQTAQSF 240
Db	181	SFNSAOSIKPLALTRSDFTAKLNQFNQDELAVYLEKFPDLBALKANRLQTAQSF 240
Qy	241	GNLVDPVYSFIRNPQNKESWASDLNODQKTVRLYLRTPEFSPQAKTILKDYKYKDETFLS 300
Db	241	GNLVDPVYSFIRNPQNKESWASDLNODQKTVRLYLRTPEFSPQAKTILKDYKYKDETFLS 300
Qy	301	SIDLKASNGTSLPANENDLKDQDLVDLLDYSDFGGQSEITITSNSQVKVPASERSLKOR 360
Db	301	SIDLKASNGTSLPANENDLKDQDLVDLLDYSDFGGQSEITITSNSQVKVPASERSLKOR 360
Qy	361	VKPKDOQKPRIEKFSLYEYDALSFYSOLOELVSKPNSIKDLVNATLARNLRSLSKYNF 420
Db	361	VKPKDOQKPRIEKFSLYEYDALSFYSOLOELVSKPNSIKDLVNATLARNLRSLSKYNF 420
Qy	421	LPDDLASHLDYFLVSKAKIKQSSITKFLFIELPIKISLSSILGDOEPNIKTLFPEKVT 480
Db	421	LPDDLASHLDYFLVSKAKIKQSSITKFLFIELPIRISLSSILGDOEPNIKTLFPEKVT 480
Qy	481	FKLDNFRDVEIEKAFGLLYPGVNEELEQARQASFEKSKKGLKEFSQOKEENSKAI 540
Db	481	FRLDNFRDVEIEKAFGLLYPGVNEELEQARQASFEKSKKGLKEFSQOKEENSKAI 540
Qy	541	NNQEGLEEDDNITIRLPENSPIQYQOENAGLGASPDKPYMIKDQVQRYYLAKSQOELI 600
Db	541	NNQEGLEEDDNITIRLPENSPIQYQOENAGLGASPDKPYMIKDQVQRYYLAKSQOELI 600
Qy	601	KAXDYTKLAKLLNRHTYNTISLRKQLFDVNPRIIPSSRDIEKAFVLDKTERKNKQWQIY 660
Db	601	KAXDYTKLAKLLNRHTYNTISLRKEQLFDVNPRIIPSSRDIEKAFVLDKTERKNKQWQIY 660
Qy	661	SSASPVQNKWSLFGYYRYLLGLDPKQTHLVLKQKAGLOFEGYENLPSDFNLEDKN 720
Db	661	SSASPVQNKWSLFGYYRYLLGLDPKQTHLVLKQKAGLOFEGYENLPSDFNLEDKN 720
Qy	721	IRIKTPLFSQDNFKLSLLDFNNYDGEIKAPEFGLPLFLPKELRRNSNSGGSQNSNP 780
Db	721	IRIKTPLFSQDNFKLSLLDFNNYDGEIKAPEFGLPLFLPKELRRNSNSGGSQNSNP 780
Qy	781	WEQEIISQFKDQNLNQDQLAQSFKIWEKIIIGDENEFDONNRLQYKLLKDQESWINKT 840
Db	781	WEQEIISQFKDQNLNQDQLAQSFKIWEKIIIGDENEFDONNRLQYKLLKDQESWINKT 840
Qy	841	RDNLWYTYLGDCLKVKPKNNLEAKFQISNLQELLTAFYTSALSNWNYQDSGAKSTI 900
Db	841	RDNLWYTYLGDCLKVKPKNNLEAKFQISNLQELLTAFYTSALSNWNYQDSGAKSTI 900
Qy	901	IFBEIAELDPKVKKEKVADYVQLKFHYAIGFDDNAGKFNQEVIRSSSRITYLTKTSGSKL 960
Db	901	IFBEIAELDPKVKKEKVADYVQLKFHYAIGFDDNAGKFNQEVIRSSSRITYLTKTSGSKL 960
Qy	961	EADTIQDLNOAVKNAPLGLQSFYLDTERFVGFQKLATSLAVQHQKEKTLPKKLNDGYT 1020
Db	961	EADTIQDLNOAVRNAPLGLQSFYLDTERFVGFQKLATSLAVQHQKEKTLPKKLNDGYT 1020
Qy	1021	LHDKLKKPVIPOISSPEKDWEGKLNQNGSQSNVNVSTFGSIIESPFSFTNQDADL 1080
Db	1021	LHDKLKKPVIPOISSPEKDWEGKLNQNGSQSNVNVSTFGSIIESPFSFTNQDADL 1080

RESULT 2
ADI28627

ID	ADI28627	standard; protein; 1878 AA.
XX		
AC	ADI28627;	
DT	XX	
DT	22-APR-2004	(first entry)
XX		
DE	Mycoplasma hyopneumoniae immunogen protein C28-MHP545.	
XX		
KW	Pneumonia; vaccine; diagnosis; C28-MHP545; immunogen; immunostimulant;	
KW	antinflammatory.	
XX		
OS	Mycoplasma hyopneumoniae.	

XX WO2004003161-A2.
XX
XX
XX PD 08-JAN-2004.
XX
XX PF 27-JUN-2003; 2003WO-US020460.
XX PR 28-JUN-2002; 2002US-0392632P.
XX PA (IOWA) UNIV IOWA STATE RES FOUND INC.
XX PA (NSWA-) NSW AGRIC.
XX FI Minion CF, Mahairas GG, Djordjevic SP;
XX DR WPI; 2004-083044/08.
XX DR N-PSDB; ADI28626.
XX PT New immunogenic Mycoplasma hyopneumoniae polypeptide, useful in eliciting
XX an immune response and in treating or preventing enzootic pneumonia.
XX PS Claim 11; SEQ ID NO 20; 81pp; English.
XX CC The present sequence is the protein sequence of C28-MPH545, an
XX immunogenic p102 paralog from Mycoplasma hyopneumoniae strain J. The
XX invention provides M. hyopneumoniae polypeptides and nucleic acids
XX ADI28608-ADI28627, including C28-MPH545, as well as vectors and host
XX cells. Compositions containing M. hyopneumoniae polypeptides and nucleic
XX acids are provided for use in methods of treating pigs to prevent
XX enzootic pneumonia. Diagnostic tests for detecting M. hyopneumoniae
XX infection in swine herds are also provided.
XX
XX SQ Sequence 1878 AA;
Query Match 97.0%; Score 9435.5; DB 8; Length 1878;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1827; Conservative 17; Mismatches 29; Indels 11; Gaps 3;
QY 1 MKNKSTLLATAAAIIGSTVFGTVGLASKVKYRGVNPQTQGISQLGLIDSVAFKPSIA 60
DB 1 MKNKSTLLATAAAIIGSTVFGTVGLASKVKYRGVNPQTQGISQLGLIDSVAFKPSIA 60
QY 61 NPTDYSQVKKALLNGKTFDPKSSFTDFVSKFDTLNNGRVLEIPKYOVISEFSP 120
DB 61 NPTDYSQVKKALLNGKTFDPKSSFTDFVSKFDTLNNGRVLEIPKYOVISEFSP 120
QY 121 DDKERFRGLGFLKEKLEGNIAQSAATKFIYLLPLDMPKAAALGQYSYIVDKNFNNLIHPL 180
DB 121 DDKERFRGLGFLKEKLEGNIAQSAATKFIYLLPLDMPKAAALGQYSYIVDKNFNNLIHPL 180
QY 181 SNFSAQSIKPLALTRSSDFIAKLNQFNNODELWVYLEKFFDLKALNIRLOTADFSFEK 240
DB 181 SNFSAQSIKPLALTRSSDFIAKLNQFNNODELWVYLEKFFDLKALNIRLOTADFSFEK 240
QY 241 GNLVDPFVYSFIRNPNQKEMASDLNQDKTVRLVLRTEFSPQATILDKYKDETFLS 300
DB 241 GNLVDPFVYSFIRNPNQKEMASDLNQDKTVRLVLRTEFSPQATILDKYKDETFLS 300
QY 301 SIDLKASNGTSLFANENDLKQDLVDLVDVDFYFGQSETITSNQVVPVPAASERSLKDR 360
DB 301 SIDLKASNGTSLFANENDLKQDLVDLVDVDFYFGQSETITSNQVVPVPAASERSLKDR 360
QY 361 VKFKDQKQPRIEKPSLYSDALSFYSQQLVSKPNSIKDLVNATLARNLRFSLGKYNF 420
DB 361 VKFKDQKQPRIEKPSLYSDALSFYSQQLVSKPNSIKDLVNATLARNLRFSLGKYNF 420
QY 421 LPDDLASHLDYFLVSKAKIKQSSITPKLFIELPIKISLKSSILGDQENIKTLPEKEVT 480
DB 421 LPDDLASHLDYFLVSKAKIKQSSITPKLFIELPIKISLKSSILGDQENIKTLPEKEVT 480
QY 481 FKLDNFRDVEIEKAFGLLYPGVNEELEQAKQASFEKEKSKGLKEFSQOKEENSXAI 540
DB 481 FKLDNFRDVEIEKAFGLLYPGVNEELEQARRSORASLEKEKAKGLKEFSQOKEENLKAI 540

QY 541 NNQGLEEDDNIITERPENSPIQYQOENAGLGASPKPYMIKDQVQRYVYLAQSIOELI 600
DB 541 NNQGLEEDDNIITERPENSPIQYQOENAGLGASPKPYMIKDQVQRYVYLAQSIOELI 600
QY 601 KAKDYTKLAKLNSRHTYINISLRLKEQLFQVNPRIIPSSRDIEKAKFVLDTEKNKYWQIY 660
DB 601 KAKDYTKLAKLNSRHTYINISLRLKEQLFQVNPRIIPSSRDIEKAKFVLDTEKNKYWQIY 660
QY 661 SSASPVPQKWSLPGYRYLLGLDPKQTHLAVKLGQKAGLQEGYENLSDFNLELKN 720
DB 661 SSASPVPQKWSLPGYRYLLGLDPKQTHLAVKLGQKAGLQEGYENLSDFNLELKN 720
QY 721 IRIKTPFSQKDNFKLSLLDFNNYDGEIKAPFGLPLFLPKELRNSSNGSGSQNSP 780
DB 721 IRIKTPFSQKDNFKLSLLDFNNYDGEIKAPFGLPLFLPKELRNSSNGSGSQNSP 780
QY 781 WSEIISQFKQDNLSNODQLAOFSTKIWEKIGDENEFQDNRLQYLLKDLQESWINKT 840
DB 781 WSEIISQFKQDNLSNODQLAOFSTKIWEKIGDENEFQDNRLQYLLKDLQESWINKT 840
QY 841 RDNLWYTLGDKLVKPKNNLEAKFROI SNLQELLTAFYTTSAALSNNWYQDSGAKSTI 900
DB 841 RDNLWYTLGDKLVKPKNNLEAKFROI SNLQELLTAFYTTSAALSNNWYQDSGAKSTI 900
QY 901 IFEBIAELDPKVKKEKVGADVYQLKFHYAIGFDDNAGKFNQEVIRSSRTIYLYKTSGSKL 960
DB 901 IFEBIAELDPKVKKEKVGADVYQLKFHYAIGFDDNAGKFNQEVIRSSRTIYLYKTSGSKL 960
QY 961 EADTIDQLNAQVNAPIGLQSFYLDTERPGVFQKLATSLAVQHKQKQKTLPKLNDGYT 1020
DB 961 EADTIDQLNAQVNAPIGLQSFYLDTERPGVFQKLATSLAVQHKQKQKTLPKLNDGYT 1020
QY 1021 LIHDKLKKPVIPQISSPEKDWPEGLKNGQSONVNVSTFGSIIESPYSTNFQEDADL 1080
DB 1021 LIHDKLKKPVIPQISSPEKDWPEGLKNGQSONVNVSTFGSIIESPYSTNFQEDADL 1080
QY 1081 DQDQDSDROGNNSLDNQEAQGLKQKLAILLGNQFIQYQQNDKEIEFEIINVEKVSLS 1140
DB 1081 DQDQDSDROGNNSLDNQEAQGLKQKLAILLGNQFIQYQQNDKEIEFEIINVEKVSLS 1140
QY 1141 PRVEFKLAKTLEDNGKTIIRVLSDETMSLI VNTTITEKPEMSAVPEVDTKWVEQYDPRTP 1200
DB 1141 PRVEFKLAKTLEDNGKTIIRVLSDETMSLI VNTTITEKPEMSAVPEVDTKWVEQYDPRTP 1200
QY 1201 LAATKTVLPKDOI PVDGSCNISKWLASTPLVHQMLRLSPVVKTIIRGLKTEQQQ 1260
DB 1201 LAATKTVLPKDOI PVDGSCNISKWLASTPLVHQMLRLSPVVKTIIRGLKTEQQQ 1260
QY 1261 QQQQQQQQQQKAVRKEELETYNPKDEFNINPLTKAHLTLNPLVNDPNYKIEDL 1320
DB 1261 QQQQQQQQQQKAVRKEELETYNPKDEFNINPLTKAHLTLNPLVNDPNYKIEDL 1320
QY 1321 KVIKNEAGDHQALFSLRANNIKRLMNTPIITADYNPPFYNNEDWRSIDKYLNNKGNVSS- 1379
DB 1321 KVIKNEAGDHQALFSLRANNIKRLMNTPIITADYNPPFYNNEDWRSIDKYLNNKGNVSS- 1379
QY 1379 KVIKNEAGDHQALFSLRANNIKRLMNTPIITADYNPPFYNNEDWRSIDKYLNNKGNVSS- 1379
DB 1379 KVIKNEAGDHQALFSLRANNIKRLMNTPIITADYNPPFYNNEDWRSIDKYLNNKGNVSS- 1379
QY 1435 LVERSIGVPMSTGLDGYIGSEQTQKDTSSSQKGFQDDFIQALGLKNTYHGLGLSIR 1495
DB 1435 LVERSIGVPMSTGLDGYIGSEQTQKDTSSSQKGFQDDFIQALGLKNTYHGLGLSIR 1495
QY 1495 IFDPGNEIAKIKDASNNKKEEKLKSYDLFKYLYNEYEKSPKIAKGTNTHPDKQEPN 1554
DB 1495 IFDPGNEIAKIKDASNNKKEEKLKSYDLFKYLYNEYEKSPKIAKGTNTHPDKQEPN 1554
QY 1554 PNQKLPENYLVNLPQKVTLYNSSDFITNLVFEPEGSDRGSGTKAQVQKQNNYA 1614
DB 1554 PNQKLPENYLVNLPQKVTLYNSSDFITNLVFEPEGSDRGSGTKAQVQKQNNYA 1614
QY 1614 DWGSAYLTFWYDKNIITNQPNVITANIADVFVKVKELEDNTEKLIAPNITQWPNISGSK 1675
DB 1614 DWGSAYLTFWYDKNIITNQPNVITANIADVFVKVKELEDNTEKLIAPNITQWPNISGSK 1675

Qy	1052	OSQNVNVTGSIIESYFSTNFQEDADLDQDQDSDRGNSLDNQEGALLKQKLAILL	1111
Db	976	NSLTKLTF---LISFLNKKDKNPKDLKADNKNDSPINPI-----IARQKLI	1026
RESULT 4			
ID	ADI28617	standard; protein; 1001 AA.	
XX	AC	ADI28617;	
XX	DT	22-APR-2004 (first entry)	
XX	DE	Mycoplasma hyopneumoniae immunogen protein C28-MHP662.	
XX	KW	Pneumonia; vaccine; diagnosis; C28-MHP662; immunogen; immunostimulant;	
XX	KW	antiinflammatory.	
XX	OS	Mycoplasma hyopneumoniae.	
XX	FH	Key Location/Qualifiers	
FT	Misc-difference 1	/note= "Encoded by TTG"	
FT	WT	WO2004003161-A2.	
XX	PN	08-JAN-2004.	
XX	PD	27-JUN-2003; 2003WO-US020460.	
XX	PF	28-JUN-2002; 2002US-0392632P.	
XX	PA	(IOWA) UNIV IOWA STATE RES FOUND INC.	
XX	PA	(NSWA-) NSW AGRIC.	
XX	PI	Minion CF, Mahairas GG, Djordjevic SP;	
XX	WPI	WPI; 2004-083044/08.	
DR	N-PSDB	ADI28616.	
XX	PT	New immunogenic Mycoplasma hyopneumoniae polypeptide, useful in eliciting	
XX	PT	an immune response and in treating or preventing enzootic pneumonia.	
XX	PS	Claim 6; SEQ ID NO 10; 81pp; English.	
XX	CC	The present sequence is the protein sequence of C28-MPH662, an	
XX	CC	immunogenic polypeptide from Mycoplasma hyopneumoniae strain 232. The	
XX	CC	invention provides M. hyopneumoniae polypeptides and nucleic acids	
XX	CC	ADI28608-ADI28627, including C28-MPH662, as well as vectors and host	
XX	CC	cells. Compositions containing M. hyopneumoniae polypeptides and nucleic	
XX	CC	acids are provided for use in methods of treating pigs to prevent	
XX	CC	enzootic pneumonia. Diagnostic tests for detecting M. hyopneumoniae	
XX	CC	infection in swine herds are also provided.	
XX	Sequence	1001 AA;	
Qy	Query Match	4.3%; Score 415; DB 8; Length 1001;	
Db	Best Local Similarity	22.4%; Pred. No. 2.2e-15;	
Db	Matches	255; Conservative 161; Mismatches 397; Indels 326; Gaps 54;	
Qy	2	KVKSTLLATAAAIIGSTVFQTV---VGLASKVKYRGVNPQTQVIGSLGLIDSVAKP	57
Db	24	KKKSTNLGRKLLTGAIVFFGIAITITPLVTVANWKIKOPRLQVQNAKLITNIQLKD	83
Qy	58	STANFTSDYQSVKCALLNGKTFDPKSSEPTDFVSKFDPLTNGRTVLEIP-----KKY	110
Db	84	EYQNGNLSTFDLKKQFNADNTKKTGIDISQF---FDYQKN-NTSLPINFATDYGNRY	139
Qy	111	QVVISFSPDDKFRFLGFLHLEKLEDGNIACS---ATKFTYLLPLDMPKAAALQGSYI	167
Db	140	KLDVDFDLQDQEQSFEIYRLVYQLPDDKKAISDLITQKVIWNY---LPDYSLANFANF	196

Qy	168	VDKNFNLLIHLPLNSFSAQSIKPLA-LTRSSDFIAKLQNFQNDLWVLEKFFDLEALK	226
Db	197	SSSKLEKRAYTNKEISLSTKKELTKLVKLEDFEKQVNAINNNEARKINKYFNLEEII	256
Qy	227	ANIRLQADFSF--EKGNIWDPFVYSFIRNPQNKQEWASD--LNQD-----QKTV-RLY	275
Db	257	AEI-LNNKEFSYLDSE-----IWNPOYQIELVRDQILGODFLAKTGOKGIYKLT	305
Qy	276	LRTERSPQ-AKTILKDYKVKDETFLLSSIDLKASNGTSLPANENDLKQDLDDVLDSDYF	334
Db	306	FYAASFENFAKTAADLN-KSKFHFGINILNN---LP-----LD-----	342
Qy	335	GGQSETITNSQVPPASERSLKDRVKFKDOOKPRIEFKFSYEDALSFYQLOELVS	394
Db	343	-----KTVAE--IKITEFSEDDYYPQINFEKLEA-----EINGWDFLNYN--NQIFA	388
Qy	395	KPNSIKDLVNNATLARNLRFSLGKYNFLFDDLASHLDYFLVFLSKAKIKQSSITKLFIEP	454
Db	389	TQNEREDFLKNLIAKIVRTPLK-KVEFENKLSGIDY-----AKFLK-----YUKLD	434
Qy	455	IKISLSSILGDQEPNI-KTLFEKVTFKLDNFRDVEIEKAFGLLYPGVNEELEQAKA	513
Db	435	IKLDANSTKLAFAKNQIVAKIFGKIILRNAE--QIVAEKNF-----SQTIEHLNRLG	485
Qy	514	RASFEKEKSKGLK-BFSQOKBENSKAINNQEGLEDDNITERLPENSPIQOQENAGLG	572
Db	486	QNDALVKQIKQTKFEF---KPETKXKIANQG-----	515
Qy	573	ASPDKPYMIKDVQNRYYLAKSQIQLIKAKQVTKLAKLSNRHTYINISL---RLKSQLP	629
Db	516	-----APKSEILALLNANKDFLKNILENGDYGYGYEFNEERLK--LL	555
Qy	630	DVNPRISSRDIEKAFVLDKTEK-----NKYM-----QIVSSASPVFQNKWSLFGYRY	679
Db	556	VHNSQLPNVEEFAPKLSVVPKEMSEGIINLWKNFKTNQEVSTFLSLAKRDISPAKYWY	615
Qy	680	LLG-----LDPKQTIHELVLKQAGLQFEGYENLPSDFNLDELKNTIRIKTFLPSQDNF	734
Db	616	DLNKKFLIDPKT-----QWPNLDQNSLFPKHLSQLKIOPP---EKKAV	656
Qy	735	KLS-----LLDPNNYDGEIKAPFGLFLFKELRNSNSGSGSONSPWEQIISQPK	790
Db	657	SLTSDFWLFLSLNDY-----LISFDYLN--	684
Qy	791	DNLSNQDLQAQFSTKIWEKIIGDNEFDQNNRLOYKLLKLOESWINKTRDNLVWYLG	850
Db	685	HSNLKNTLDL-----IKTESAFNRDFVEH--IRELAKS-----	716
Qy	851	DKLVKPKNNLEAKFRQ-IGNLOELLTAPYTSAAALSNNWNYQDSGAKSTIIFBEIAB-L	908
Db	717	-----IKPKDFIQEKGKNPITNLSEFLVAFVELI-----YSKQDQ---LLAESLGQNL	761
Qy	909	DPKVK-----EKGADV-----VOLKFAHVAIGFDDNAGKNOEIVIRSSS	947
Db	762	DYKIQFELEPISLNAVVSQSKTNPNNNLRLNRLKWKYKIGSDVQNGNLIQVIYQTK	821
Qy	948	RTIYLTSGSKLEADTIDOLNQAVKNAPLGQSLQFYLDTFRFGVQKLSLATSLAVQHKQ--	1005
Db	822	ETLQVNNNNKLLSEDEVEKLEIATNPPSADQIIFLKEDY-----TQVDSIKQVI	874
Qy	1006	KEKTLPKKLNNDGYTLIHDKLKPKVPIQISBPEDKMFEGKLNQNG-----OSQNVNVS	1059
Db	875	KTENTPVKIDN-----QIKNLPFSQFFENYPDYGVFIYIKTSKNLESS	917
RESULT 5			
ID	ABU48941	standard; protein; 4688 AA.	
XX	AC	ABU48941;	
XX	DT	19-JUN-2003 (first entry)	
XX	XX		

Db 483 -QOYL-----LITNSVNDKSGPLR-----KEEIQFINIQEDDSTITESDSQKVT 530
Qy 217 EKFFDLKALNIRLQADPFSFK--GNLVDPPFYSPIRPNQKQEWASDLNODQKTVRL 274
Db 531 ERLVEFKNI---IQOEKNAELLKVVNADKLESKEKSKQSLOKTESVTNEAEAIL 587
Qy 275 YLRTEFSPQAKTILKOYKDE--TFUSSIDLKASNGT--SLFANENDLKQDLDVLLDV 330
Db 588 TLASEKMDLESRIEELQELKELKTSVPNEDASVNTIKQLTETKRDLESQVDLQTRI 647
Qy 331 SDYFGGSEFIT--SNSQVKVPASERSLKDVRFEKQDQKPRIEKSLFEYDALSFSQL 389
Db 648 SQITRESTENMULNKIEQIDLYSKSDISIKLGEKSSRIABERFKL----- 695
Qy 390 QELVSKPNSIKDVLNATLARNLAFSLGKYNFLFDLASHLDYFLVSKAKIKOSSITKIL 449
Db 696 -----LSNLTDLTKAEN-----DQLRKFDY---LQNTILKQDSKTHET 731
Qy 450 FIELPIKISLSSILGQOEENIKTLFEKVTFTKLDNPRDVEIEK-----AFG 496
Db 732 LNEY-VSCSKSLSVETELLNLKE--EQKLRVHLEKNLQOELNKLSPKDSLRIMVTQLQ 788
Qy 497 LLYPGVNEELEQARKA--ORASFEEKSKKGLKFPQOKEENKAINNQEGLEEDDNIATER 555
Db 789 TLQKEREDLLEETRKSCOKIDELEDALEKETSQKHIIKOL-----BEDN----- 837
Qy 556 LPENSPIOYOQENAGLQASDPKPMIKDVQONQRYLAKSQ--IQELIKAKDYTKLAKLLS 613
Db 838 ---NSNIEWQ--NKIEALKDYESVITSVDKQTDIEKLQYKVKSEKEIEEDKI----- 888
Qy 614 NRHTYNTSLBKQLPQVNPRISSRDIEKAKFVLDK--TERKNYQWQYSSASVPQNKW 671
Db 889 RLHTYNY--MDETINDDSL--KELEKSKINLTDAYSQIKYKDYVETTSQSLOQTN 941
Qy 672 SLFGYRYLLGLDP-----KQTHIELVLGQKAGLQFEGVENLPSPDFNLEDLKNIR 722
Db 942 S-----KLDESFKDFNQIKNLTDKTSLEDKISLKEQMFNLNDELIDQ----- 986
Qy 723 IKTPLFSQKDNFK--LSLLDFNNYDGEIKAPFGLPLFLPKELRLRNSNGSGSQNSNP 780
Db 987 -KKGMEKADFKKRISILQNNKNEVEAVKSEYKSLKIQNDLDDQOTIYANTAQN--- 1042
Qy 781 WEQEI-----ISQFKDQ-----NLSNQDLAQFSTKIWEKIGDENEFDQ 820
Db 1043 YEQELQKHADVSTISLREQLHTYKGVKQVTLNLS-RDQL-----EN----- 1083
Qy 821 NNRLQYLLKLDQESMINKTRDNLWYTLGDKLVKPKNLEAKFRQISNLQELL---TA 877
Db 1084 -----ALKENKSSWQKESLL-----EQDLS-----NSRIEDLSQNKLLYDQIQ 1125
Qy 878 FYTSAALSNNWNYQDSGAKSTIIF-----EEIAELDPKV----- 912
Db 1126 IYTAADKEVN--NSTNGPGLNNILITLRRERDILDTKVYVAERDAKMLRQKISLMDVELQD 1184
Qy 913 -KEKVGADVQLKPHVAI--GFDNAGKFNQ--EVISSRTI--YLTSGKSKLEADT-I 965
Db 1185 ARTKLDNSRVEKENHSSIIQQHDDIMEKLNQLNLLRESNITLRENNNNKKELQSEL 1244
Qy 966 DOLNOAVKNAPLGL-----QSPYLQTERFGVFOKLATSLAVCHQKKTLPK 1012
Db 1245 DKLQNV--APIESELTKYSMQEKQELKLAKEEVHWRKKSQDILEKHEQSSSDYE 1302
Qy 1013 KLANDGYTLHD-----KLKFPVIFIQSSPEKDWFEGLKNQGSQNV 1056
Db 1303 KLESIEENLKELENKEROGAEEKFNLRQQAQERLKS-----KLSQDSITEQV 1354
Qy 1057 N-VSTFGSIIIESYFTSNFOEDADLQDQDSDRQGNNSLD----- 1096
Db 1355 NSURDAKNVLENSLSBAN-----ARIEQLQNAVQAGNQLQEAIRKLQEDAERLOAK 1410
Qy 1097 -----NOEAGLLK-----QKLAILLGNQ----- 1114

Db 1411 LBESTTSYESTINGLNEEBITTLKBEIEKORQIQOQLQATSAEQNDLSNIVESNKKSFEE 1470
Qy 1115 -FIQYYQNDKEIEFEI-----INVEKV-----SELSFRVEFKLAKTLEDNGK 1156
Db 1471 DKIKFIKTEQVNEKILEAQERLQPSNINMEEIKKKWSEHEQEVSKIREABEALKK 1530
Qy 1157 TIRVLSDETMSLIVNTTIE-----KTPMSAVPEVFTDKVQVQDPTPLA 1202
Db 1531 RIRLPTEEKINKIIRKKEELEKEFEKVEERIKSMEQSGEIDVVLKQLEAKVQEKQKE 1590
Qy 1203 AKTKFVLKFKDQI-PVDGSGNIS---DKWLASIPLVIHQQLMLRLSPVVKTIRELGLKTE 1257
Db 1591 LENEYNKQLEELKQVPHSSHISDDERDKLRAEISRLREEFNNELOAIKK-----KSF 1644
Qy 1258 QOQOQOQOQOQOQOQKAVRKEEEL-ETYN-----PKDEFNINLPLTKAHLRLTSLNVN 1311
Db 1645 DEGKQQAQMMKTTLLERKLAKMESQLSETKQSAESPSPKSVNNVQNPLGLPRKIEEN--SN 1702
Qy 1312 DP-NYKIEDLVKINEAGDHQAFSLRANNIKRLMNTPTITPADYNPFFYFNEDWRSIDKY 1370
Db 1703 SPFNPLLSGKLLKLNLSKSSSGFN-----PFTSPSPNKLQNDNDKR---ES 1747
Qy 1371 LANKGNVSSHQQAAGNQGSLI---QRLNKNIKPETFTPALIALKDRNNTNLSNYSK 1427
Db 1748 LANKDPPHLEPSPFNIPASRGLISSSTLSTDTNDEBLTSNNPAQKDSNRNVQSEED- 1806
Qy 1428 IIMIKPKYLVERSIGVPMSTGLDGYIGSQTKDGTSSSSQKGFQDQDFIQALGLKNTEYH 1487
Db 1807 -----TEKKKEGEP-----VKRGEAIBEQTKSNKRPIDEVGELKNDE-- 1843
Qy 1488 GKLGLSIRIFDPCNELAKIKDASNKKGBEKLKSKVDLFGNYLYNEYKKSFKPI 1539
Db 1844 -----DDTTENINESKTIKTED-----EEKETDKV 1869
RESULT 7
ADK64380
ID ADK64380 standard; protein; 1875 AA.
XX
AC ADK64380;
DT 06-MAY-2004 (first entry)
XX
DE Disease treating protein complex-derived protein #1185.
XX protein complex; drug target; diagnosis.
XX Unidentified.
XX EP1338608-A2.
XX 27-AUG-2003.
XX 20-DEC-2002; 2002EP-00102902..
XX 20-DEC-2001; 2001EP-00130253.
PA (CELL-) CELLZOME AG.
PI Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;
PI Marzioch M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;
PI Michon A, Leutwein C, Rick J;
XX WPI; 2003-638460/61.
DR N-PSDB; ADK64381.
XX
PT New proteins and protein complexes from eukaryotes, useful as targets in
PT drug screening, or in diagnosing or screening for the presence of a
PT disease or disorder, or a predisposition for developing a disease or
PT disorder in a subject.
XX
PS Disclosure; SEQ ID NO 2369; 13pp; English.

CC The invention relates to novel protein complexes comprising a first and a
CC second protein, or its derivative, fragment, homologue or variant. The
CC proteins are selected from given protein complexes, which are not defined
CC in the specification. The variants are encoded by nucleic acids that
CC hybridize to the nucleic acids encoding the proteins under low stringency
CC conditions. The protein complexes are useful as targets for an active
CC agent of a pharmaceutical. These protein complexes are particularly
CC useful as drugs targets for the treatment or preventing of a disease or
CC disorder. The complexes and methods above are useful in diagnosing or
CC screening for the presence of a disease or disorder or a predisposition
CC for developing a disease or disorder in a subject. These are also useful
CC in screening for a drug for treatment or prevention of a disease or
CC disorder. The molecule that modulates the amount, activity or protein
CC components of the complex is useful for the manufacture of a medicament
CC for the treatment or prevention of a disease or disorder. This sequence
CC corresponds to a protein of the invention. (Note: the sequence data for
CC this patent did not form part of the printed specification but was
CC obtained from the EPO in electronic format).

XX Sequence 1875 AA;

Query Match 3.1%; Score 305.5; DB 7; Length 1875;
Best Local Similarity 19.6%; Pred. No. 1.6e-08;
Matches 340; Conservative 271; Mismatches 627; Indels 494; Gaps 75;
60 ANFTSDYQSVKALLNGK-----TF-----DPKSEFTDFVSKPDL-----TNGR 101
Db AKSSDFFLKQIKKRTKEHLQNETFIVELEHKVPINSGFKERTDMLNELNAA 439
Qy TVLETPKYQVVISFSPEDDKERFLGFHLEKLEDGNIASATKFIYLLPLDMPKAAL 161
Db LLE-----HTSNEKNAKVELNAKQKLVCECNDLQTLK-----QRLDLCR--- 482
Qy 162 GQSYIVDKNFNNLIHPLNFSQSIKPLALTRSSDFIAKLQ-----FNNDELWYL 216
Db 483 -QIQVL-----LITNSVNSDKPLR-----KBEIQIQMIEDDSTIESDSQVVT 530
Qy 217 EKFPDLEALKANIRLOTADPFEK--GNLVPFVYFTRNFQNKQEWASDLNQOKTVRL 274
Db 531 ERLVEFKNI---IQLEKNAELLKVVRLNADKLESKEKSKQSLQKIESETVNEAKEAI 587
Qy 275 YLRTEFSQAQKILDKYKDR--TFLSSIDLKASNGT--SLFANENDLKQDLVDLDV 330
Db 588 TLKSEKMDLESRIEQLQLELELKTSPVEDASYNVTIKQETKRLDESQVQLQTRI 647
Qy 331 SDYFGQSEITIT--SNSQVKVPASERSLDRYKFKKQDQKPRIEKFSLYEYDALSFYSOL 389
Db 648 SQITRESTENMSLLNKEIQDLVDSKSDISIKLGKSSRIILAEERFKL----- 695
Qy 390 QBLVKPNSIKDLVNATLARNRFLSGKYNFLFDDLASHLDYFVFLVAKIKQSSITKYL 449
Db 696 -----LSNTLDLTKAEN-----DQLRKFDY---LQNTILKQDSKTHET 731
Qy 450 FIELPKISLKSILGQDQEPNLTFLFEKVTFKLONFRDVELEK-----AFG 496
Db 732 LNEY--VSKSKLSIVETELNLKE--EQKLRVHLEKLNKQELNKLSPKQSLRIMVTLQ 788
Qy 497 LLYPGVNELEOAKA--ORASPEKSKKGLKEFSQOKEENSANNOEGLEEDDNITER 555
Db 789 TLQEREDLEETRSCQKKIDLELDELALSELKETSQDHHIKQL-----EEDN--- 837
Qy 556 LPENSPIOYQOENAGLGASDPKPYMIKVONQRYVLAKSQ--IQELIKAKYTKLAKLLS 613
Db 838 ---NSNIEWYQ--NKIEALKKOYSEVITSVDSQTDIEKLQYKVKLSLEKIEEDKI----- 888
Qy 614 NHTYNIISRLKEQLFDVNPRLPSSRDIKAKFVLDK--TEKNKYQIYSSASPIYFQNKW 671
Db 889 RLHTYNV---MOBTINDSLR-----KELEKSKINLTDAYSQIKYKDLYETTSSQSLQQTN 941
Qy 672 SLFGYRYLLGLDP-----KQTIHELVLKLGQKAGLQFEGYENLPDFNLELKNIR 722
Db 942 S-----KLDESFKDFNTQIKNLTDEKTSLEDKISLLEKQMFNINLELDLQ----- 986

Qy 723 IKTPLFQKDNFK--LSLLDFNNYDGEIKAPBFGPLFLPKELRRNSSSGSQNSNP 780
Db 987 -KKGMEKADFKRISILQNNNEKVAEYKSEVESKLSQNDLDDQOTIVANTANN--- 1042
Qy 781 WEQEI-----ISQFKDQ-----NLSNQDLAQFSTKIWEKIIGENEFQ 820
Db 1043 YEQELQKADVSKTISELRQLHTYKGVKTNLS--RDQL-----EN----- 1083
Qy 821 NNRLOYKLLDQESWINKTRDNLWYTLGDKLVKPKONLEAKFRQISNLQELL---TA 877
Db 1084 -----ALKENKSSQKESL-----EQDLIS-----NSRIEDLSQNKLLYDQIQ 1125
Qy 878 FYTSAALSNNWYQDSGAKSTIIF-----EETAEALDPKV----- 912
Db 1126 IYTAADKEVN--NSTNGPLNLLITLRRERDILDTKVVAERDAKRLQKISLMDVELQD 1184
Qy 913 -KKBGVADVQLKPHVAI--GFDNAGKFNQ--EVRSSRTI--YLTSGSKSLEADT-I 965
Db 1185 ARTKLONSRVEKENHSGSIIQQHDDIMEKLNQLNLLRESNITLNELENNNNKKELQSEL 1244
Qy 966 DQNLQAVKNAPLGL-----QSFVLDTERFGVFQKLATSIAVOHKQKEKTLPK 1012
Db 1245 DKLQNV--APISELTALKYSMOEKEQELKAKEEVRWKKRSQDILEKHEQLSSSDYE 1302
Qy 1013 KLNNDGYTLIHD-----KLKCPVIPQISSSPKDWFEGLKLNQNGSQNV 1056
Db 1303 KLSSEIENLEKEENKEROQAEAEKFNLRQQAERLKTSS-----KLSQSLTEQV 1354
Qy 1057 N-VSTFGSIIIESPYFTNFQEDADLDQDQDSDSQNGNSLD----- 1096
Db 1355 NSLRDAKNVLENSLEAN--ARIEBLQNAKVAQGNNOLEAIRKLOEADKASRELOAK 1410
Qy 1097 -----NOEAGLLK-----OKLAILLQNG----- 1114
Db 1411 LEESTTSYESTINGNEEITTLKEEIEKQRIQOOLQATSANQNDLSNIVESMKKSFE 1470
Qy 1115 -FIQYQONDKIEIFEI-----INVEKV-----SELSFRVEFKLAKTLEDNGK 1156
Db 1471 DKIFEIKTEQVNEKILEAQERLNQPSNINWEEIKKKWESEHEQVQSQKIREAEALKK 1530
Qy 1157 TIRVLSDETWSLIYVNTIE-----KTPEMSAPVEPFDTKWVQYDPTPLA 1202
Db 1531 RIRLPTBEKINKIIRKKEELEKEFEKVEERIISKMEQSGEIDVVLKQLEAKVQEKQKE 1590
Qy 1203 AKTFVLKFKDQI--PVDGSGNIS--DKWLASPLVHQMLRLSPVVKTIRELGLKTE 1257
Db 1591 LENEYKKLOBELKDVPHSSHISDDERDKLRAEIESLRREEFNELQAIKK-----KSF 1644
Qy 1258 QOQOQOQOQOQOQKQKAVRKEBEL--ETYN-----PKDEFNINPLTKAHLTLNLYNN 1311
Db 1645 DEGQOQAMKTTILERKLAKMESQLSETKQAEPPKSVNNVQNPPLGLPKRIEEN--SN 1702
Qy 1312 DP--NYKIEDLVIRKNEAGHOLAFSLRANNIKRLMNTPTITADYNPFYNNEDWRSIDKY 1370
Db 1703 SPFPNLLSGEKLKLNKSSSGGFN-----PFTSPSPNKHQLQNDNDKR---ES 1747
Qy 1371 LNNKGNVSHHQOQAGNGQSGLI---QRLANKIKPETFTTALLIALKDRNNTNLSNYS 1427
Db 1748 LANKTDPPTHELPFNIPASRGLISSSSTLTDTNDELTNNPAQKSSNNRNOSEED- 1806
Qy 1428 IIMIKPKYLVRSIGVPWSTGLDGYIGSEQTKDGTSSSSQKGFQDQFQIALGLKNTRYH 1487
Db 1807 -----TEKKKEGP-----VKRGEAIEEQTKSNKRPIDEVELKANDE-- 1843
Qy 1488 GKLGISIRIPDGNELAKIKDASNNKKGEEKLLKSYDLFKNYLVNEVEKSPKI 1539
Db 1844 -----DDTTENINESKKIKTED-----EBEKETDKV 1869

RESULT 8
ADS43855
ID ADS43855 standard; protein; 1875 AA.
XX

AC	ADS43855;	Db	380	AKSSDFIFLKKQLIKERTKEHLQNIETPTVELEHKKVPIINSFKERTDMLNENLNAA	439
XX		Qy	102	TVLEIPKQYQVWISSEPEDDKERFRLGPHLKEKLEDGNIASATKFIYLLPLDMPKAAL	161
XX	02-DEC-2004 (first entry)	Db	440	LLLE-----HTSNEKNAVKVELNAKQKVECDNLQTLTK-----QRDLCLCR---	482
XX	Bacterial polypeptide #22285.	Qy	162	GQSYIVDQNFNNLIHPLSNFSAQSIKPLALTRSSDFIAKLNQ-----FNNODELWVYL	216
KW	Recombinant DNA construct; transformed plant; improved plant property;	Db	483	-QIYVL-----LITNSVSNDSKGPLR-----KEEQFTQINQMEDDSTITESDQKVVT	530
KW	cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;	Qy	217	EXFFDLEALKANIRLOTADFSFEK--GNLVDPVYSFIRPNQOKEMASDNLNDQKTVRL	274
KW	pathogen tolerance; pest tolerance; plant disease resistance;	Db	531	ERLVEFKNI---IQLEKVAELLKVVRLADKLESEKESKQSLQKIESETVNEAKAII	587
KW	cell cycle pathway modification; plant growth regulator;	Qy	275	YLRTFSPQAKTILKDYKYKDE--TFLSSIDLKASNGT--SLFANENDLKQDLVDLLDV	330
KW	homologous recombination; seed oil yield; protein yield; carbohydrate;	Db	588	TLKSEKMDLESRIEELQKELEBELKTSVPNEDASYSNVTIKQLTETKRDLESQVODLQTRI	647
KW	nitrogen; phosphorus; photosynthesis; lignin; galactomannan;	Qy	331	SDYFGQSETIT--SNSQVFPVPSERSLKDRVKFKDQKQKPRIEKFSLEYDALSFYSQL	389
XX	bacterial polypeptide.	Db	648	SOITRESTENMLLAKIEIQDLYDSKSDISIKLGKESKSRILABERFKL-----	695
OS	Bacteria.	Qy	390	QELVSKPNSIKDLVNATLARNLRFSLGKYNFLDFDDLASHLDYVFLVSKAKIKOSSITKKL	449
XX	US2003233675-A1.	Db	696	-----LSTNLDLTKAEN-----DQLRKFDY---LQNTILKQDSKTHET	731
XX	18-DEC-2003.	Qy	450	FIELPIKISLKSIIQDQEPNITLFEKVTTFKLNFRDVEIEK-----AFG	496
XX	20-FEB-2003; 2003US-00369493.	Db	732	LNEY-VSCSKLSIVETELLNLKE--EQKLRVHLEKLNQELNKLSPKDSLIMVTLQ	788
XX	21-FEB-2002; 2002US-0360039P.	Qy	497	LLYPGVNEBLEARKA-QRASFEKSKGLKEFQOKEENSKAINNQGLEBDDNITER	555
XX	(CAOY/) CAO Y.	Db	789	TLQKEREDELEETRSKQCKKIDELDALSELKETSQKDHIIKQL-----EEDN----	837
XX	(HINK/) HINKLE G J.	Qy	556	LPENSPIOYQOQNAGLSPDKPYMIKDVONORYVLAKSQ--IOELIKAKDYTKLAKLLS	613
XX	(SLAT/) SLATER S C.	Db	838	-----NNNIEWQ-NKIEALKDYVESITSDVSQKTIEKLOYKVKSEKIEBEDKI-----	888
XX	(CHEN/) CHEN X.	Qy	614	NRHTYNISLRLKEQLFDVNPRISSRDIEKAFVLDK--TEKNKYWQIYSSASPVFQNKW	671
XX	(GOLD/) GOLDMAN B S.	Db	889	RLHTYVNV---MDETINDDSLRL---KELEKSKINLTDAYSQIKYKDYVETTSQSLOQTN	941
PI	Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;	Qy	672	SIFGYRYLLGLDP-----KQTIHELKVLGQKAGLOFEGYENLPDFFNLEDLKNIR	722
XX	WPI; 2004-061375/06.	Db	942	S-----KLDESFKDFTNOIKNLTDKTSLEDKISLLEKQMFNLNLEDLQ-----	986
XX	New recombinant DNA construct comprising a promoter positioned to provide	Qy	723	IKTPLFSQKDNFK--LSLIDFNMYDGEIKAPEFGLPLFLPKELRRNSNSGGSQNSNP	780
XX	for expression of a polynucleotide encoding a polypeptide from a	Db	987	-KKGMEKADFKKRISILONNNKEVAEYKSEKLSKIQNDLDOQTIYANTAQNN---	1042
XX	microbial source, useful for producing plants with improved properties.	Qy	781	WEQEI-----ISQFKDQ-----NLSNQDLAQFSTKIWEKIIGDENEFQ	820
XX	Claim 1; SEQ ID NO 22285; 122pp; English.	Db	1043	YEQELQKHADVSKTISELREQLHTYKGVKTLNLS-RDQL-----EN-----	1083
XX	The invention relates to a recombinant DNA construct comprising a	Qy	821	NNRLQYLLKDLQESWINKTRONLWYTLGDKLVKPKNNLEAKPQISNLQELL---TA	877
XX	promoter functional in a plant cell, where the promoter is positioned to	Db	1084	-----ALKENEKGSQKESLL-----BQDLDS-----NSRIEDLSQNKLLYDQIQ	1125
XX	provide for expression of a polynucleotide encoding a polypeptide from a	Qy	878	FYTSAAALSNNWNYQDSGAKSTIIF-----EETAELDPKV-----	912
XX	microbial source. The invention also relates to a transformed plant	Db	1126	ITYAADKEVN-NSTNGPGLNNLILITLRERDILDTKYTVVAERDAKMLRQKISLMDVELQD	1184
XX	comprising the recombinant DNA construct and a method of producing a	Qy	913	-KEKVGADYVQKPHYAI--GFDDNAGKPNQ-EVIRSSRTI--YLTSGKSKLEADT-I	965
XX	transformed plant having an improved property. The plant is a crop plant	Db	1185	ARTKLDNSRVEKENHSSIIQQHDDIMEKLNQNLNLLRESNITLRNELENNNNKKELQSEL	1244
XX	such as maize or soybean. The method of producing a transformed plant	Qy	966	DOLNQAQKAPLGL-----QSFYLDTERFGVFOKLATSLAVQHKKQKETLPK	1012
XX	having an improved property comprises transforming a plant with the	Db	1245	DKLKQNV--APIESELTKYMSQKQELKLAKEVHVHKKRSQDILEKHQELSSDYE	1302
XX	recombinant DNA construct and growing the transformed plant, where the	Qy	1013	KLNNQGYTLIHD-----KLKFPVTPQISSPEKOWFEGKLNQNGSQNV	1056
XX	polynucleotide or polypeptide is useful for improving plant properties.	Db	1303	KLSEIENLKELENNKERQGAEBEKNFLRQAQERLATS-----KLSQDSLTEQV	1354
XX	The recombinant DNA construct is useful for producing plants with				
XX	improved plant properties, e.g. improved cold, heat or drought tolerance,				
XX	tolerance to herbicides, extreme osmotic conditions, pathogens or pests,				
XX	increased resistance to plant disease, better growth rate by modification				
XX	of the cell cycle pathway with plant growth regulators, increased rate of				
XX	homologous recombination, modified seed oil or protein yield and/or				
XX	content, improved yield by modification of carbohydrate, nitrogen or				
XX	phosphorus use and/or uptake, by modification of photosynthesis or by				
XX	providing improved plant growth and development under at least one stress				
XX	condition, improved lignin production or improved galactomannan				
XX	production. This sequence represents a bacterial polypeptide used in the				
XX	scope of the invention. Note: The sequence data for this patent did not				
XX	form part of the printed specification but was obtained in electronic				
XX	format from USPTO at seqdata.uspto.gov/sequence.html.				
SQ	Sequence 1875 AA;				
Query Match	3.1%; Score 305.5; DB 8; Length 1875;				
Best Local Similarity	19.6%; Pred. No. 1.6e-08;				
Matches	340; Conservative 271; Mismatches 627; Indels 494; Gaps 75;				
Qy	60 ANFTSDYQSVKALLNGK-----TF-----DPKSEFTDFVSKPDFL---TNNGR 101				

Db 1214 TERDHLRGYIREIE---ATGL---QTKBELKIAH-----IHLKHEQETIDELRRSVEK 1261
Qy 537 SKAINNOEGLEEDNITERLPENSPYOQOENAGLQASDPKPYMIKDQVQRYYLAKSQI 596
Db 1262 TAQIINTQDLEKSH---KLQBEIPVLHBEQEL-----LPNVKVKSETQETMNELELL 1311
Qy 597 QELIKADYTKLAKLLSNRHTYINISLRLKEQLFDVNPRIPS-SRDIEKAKAFVLDKTEKNK 655
Db 1312 TEOSTTKDSTTLARIEMER-----LRLNEKFQESQBEIKSLTKERDNLKTIKEALEV-K 1364
Qy 656 YMOIYSSASPVFQNKWSLFGYRYLLGLDPKQITHE-LVKLGQKAGIOPFGYENLPSDFN 714
Db 1365 HDQU-----KEHIRETLAKIQESQSKQ-----BQSLN 1391
Qy 715 LEDLKNIRIKTPLFSQKDNFKLSLLDFNNYDGEIKAPEGLFLFLPKELRRNSSGGS 774
Db 1392 MKEKDN--ETTKIVSEMEQFKPK-----DSALLRIBIEM-LGLSKRLQESH-----1435
Qy 775 QNSNSPWEQEIISQFDQONLSNQDLAQFSTKIWE---KIIGDENFQDNRLQYKLLKD 831
Db 1436 -----EMKSAKEDDLQRLQEVLSQESDQLKENIKETIVAKHLETEEBELKVAHCLKE 1488
Qy 832 LQSSWINKTRDNL-----YWTYLGDKLVKPKNNLEAK---PROISNLQELL 875
Db 1489 -QSETINELRVNLSEKETETISTQKLEAINDLQNKIQIYEKBSQNLKIQISEVQENV 1547
Qy 876 TAF-----YTSAALSNNWNYQDSGAKSTIIFIEIAELDPKVKKEGVADV 920
Db 1548 NELKQFKHRKAKDSALQSTESKMLELTNRLQSSQBEIQIMIKESEM-KRVOEALQIER 1606
Qy 921 YQLKHYAIGFDDNAGKFNQEVIRSSRTI-YLKTSGSKLEAD--TIDOLNOAVKNAPL 977
Db 1607 DQLK-----ENTKEIVAKMESQEKBYQFLKMTAVNSTQKMCIEIHLKEQFETQKL 1658
Qy 978 GLOSFLYDTERFQVKLATSLAVQH-----KQKEKTLPKLNNQGYTILHDKLKK 1028
Db 1659 NLEN--JETENIRLTQLHENLEMSVTKERDLSRVETL--KVERD--QLKENLRE 1711
Qy 1029 PVIPOISSSPKQWFEGLKNGQSNVNVSTFGSIIIESPYFTNFDQADLDQDQ----1084
Db 1712 TITRDLEKQBEELKIVHMLKHEH--QETIDKLRGIVSEKTNESNNQKDLHSNDALKAQ 1768
Qy 1085 -----QDSSROGNSLNDQ-----AGLLKQKLAILL-----GNQFIQYQNDKEI 1126
Db 1769 DLKIQBELRIAHMLHKEQBTIDKLRGIVSEKTKLSNMQKOLENSAKLQKIEQLKAN 1828
Qy 1127 EFBIIINV-----EKVSLSFRVEFKLAKTLEDNGKTRVLSDETMSLIVNTTIEKTP 1178
Db 1829 EHQLITLKVDVNETQKVSEME-----QLKKQIKDQSLTSLKLEIENLNL-AQELHENLE 1882
Qy 1179 EMSAVPE-----VFDTRKWEQVDPRTPLAAKTKFVLKPKQIIPVDGSGNISDKWLASI 1231
Db 1883 EMKSVMKERDNLRRVETLKLERDQLKESLQETKARDLEIQELK-----TCSEK-----1932
Qy 1232 PLVTHQOMLRLSPVKVTIRELGLKTEQOQQOQQOQQOQQOQKAVKAKEBELETYNPKDEF 1291
Db 1933 ---ISEKTIQISDIQK-----DLDKGKDELQKIQBELQKLEQLLQKVEDVNNSHKK--I 1982
Qy 1292 NILNPLTKAHLRLTSLNLVNDPNVKIEDLKVINEAGDHQALFSLRANNIKRLMNTPIIF 1351
Db 1983 NEMEQLKKQF-----EPNY-----LCKCEMDNFQTKKLH-----ESLEEIRIVA 2022
Qy 1352 ADYNPPFYNNEDWRSIDKYLNNKN--VSSHQOQAAGGNGGSLIORLNNKHIPKFTTFA 1409
Db 2023 KE-----RDELRRIKESLKMEROQFIATLEMIARDRO-----NHQVKPEK-----2063
Qy 1410 LIALKDBNNTNLSYDKIMIKPKVILVERSIGVPHSTGLDGYIGSEQTKDGTSSSSQOK 1469
Db 2064 -RLLSGQOQLHMSLEKCSRIKE--LLKK-----YSEMDDHV-----2098
Qy 1470 GFQDQDFQALGL---KNTYEHG-----KLGLSIRIFDPGNELAKIKDASNKKEEKLKS 1521
Db 2099 ----ECLNRLSLDLEKIEFHRIWKLYVLSY-----VTKIKEE-----QHCINK 2141

Qy 1522 YDLFKNYLNEYKKSPIAKGWTNIHPDQKEYPNPNQKLPENYLNVLNQPWKVTL-----1577
Db 2142 FEM--DFIDEVEKQKELLIK---IQHLQDQCDVPSRELRD-----LKLQNMDLHIEEIL 2191
Qy 1578 --YNSSDF-----ITN-----LFEVEPSGDSRSGTKLKQVIOK-----QVNNNY 1614
Db 2192 KDFSESEFPIKTEFQQVLSNRKEMTQFLFEELWLNTRFDIEKLNKGIQKENDRICQVNN--2249
Qy 1615 ADWGSAYLTFWYDKNI-ITNQPNVI---TANIADVPFKDVKEL-EDNTKLIAPNITQWWP 1669
Db 2250 -----FFNRIIAINVESTETEERSATISKWEQDLKSLKEKNEKLFKNYQTLKTS 2300
Qy 1670 NTSGSKEKPYKPTVFGNWNENSSNMNSQAQPTTWEKIRE 1709
Db 2301 LASGAQ---VNPT-----TQDNKNPHVTSRATQLTTEKIRE 2333
RESULT 10
AAM39097
ID AAM39097 standard; protein; 2663 AA.
XX AAM39097;
AC AAM39097;
XX 22-OCT-2001 (first entry)
DT 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 2242.
DE Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemocastic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX Homo sapiens.
OS Homo sapiens.
XX WO200153312-A1.
FN 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US034263.
PF 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao Q;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR N-PSDB; AAI58253.
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX Example 4; SEQ ID NO 2242; 10078pp; English.
PS The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAW42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and

Db 1214 TERDHLRGYIREIB---ATGL---QTKBELKIAH-----IHUKHEQSTIDEARRSVSEK 1261
Qy 537 SKAINNOGLEEDDNIITERIPENSPIYOQENAGLGASDPKPMIKDVQNRYYLAKSQI 596
Db 1262 TAQIINTQDLEKSH---KLQEBIPVLHSEQL-----LPNVKKVSETQETMNELELL 1311
Qy 597 QELIKADYTKLAKLNSRHTYINISRLKEQLFDVNPRIPS-SRDIEKAKFVLDKTEKNK 655
Db 1312 TEOSTTDDSTTLARIEMER-----LRNBEKFQESQBEIKSLTKERDNLNLTKEALEV-K 1364
Qy 656 YWQIYSSASPQNKWSLFGVYRYLLGLDPKQITHE-LVKLGQKAGIQFQGENLPSPDEN 714
Db 1365 HDQL-----KEHIRETLAKIQESQSKQ-----EQLSLN 1391
Qy 715 LEDLKNIRIKTPLFSQDNFKLSLLDPNNYDGEIKAPBFLPLFKELRRNSSNGGS 774
Db 1392 MKEKDN--ETTKIVSEMEQFKPK-----DSALLRIEIM-LGLSKRLQESH-----1435
Qy 775 QNSNSPWEQEIISQFKDNLNSNQDLAQFSTKIWE-----KIIGDENEPDQNNRLOYKLLKD 831
Db 1436 -----EMKSAKEKDDLRQLOEVLOSBSQDLKENIKEIVAKHLETEEBELKVAHCCLE 1488
Qy 832 LOESWINKTRDNL-----YWTYLGDKLVKPKNNLEAK-----PROJSLNQLLEL 875
Db 1489 -QETINELRNLSSEKETEISTQKLEAINDKLQNKIQEIYKEBEQLNLIKQISEVOENV 1547
Qy 876 TAF-----YTSAAALSNWNYQDSGAKSTIIFBIEABLDPKVKEKGVADY 920
Db 1548 NELQKQKEHRKAKDSALQIESKMLELTNRLOESQEBIQIMIKEBEM-KRVQEAQIER 1606
Qy 921 YQLKFHYAIGFDDNAGFNQEVNTRSSRTI-YLKTSGSKLEAD--TIDOLNQAQVKNAPL 977
Db 1607 DOLK-----ENTKEIVAKMESQEKYQFLKMTAVNETQKMCBEIEHLKEQFETQKL 1658
Qy 978 GLQSFYLDTRFRFGVFOKLATSLAVQH-----KQEKTLPKKLNDGYTILHDKLKK 1028
Db 1659 NLEN--JETENIRITQILHENLEBMSVTVKERRDLRSVESTL--KVERD---QKENLRE 1711
Qy 1029 PVIPISSSPKQWFKGLNQGSQNVNVTSGSIIESPYFTNFQEDADLDQDQ---1084
Db 1712 TITRDLEKQEBELKIVHMLKHEH---QETIDKLRGIVSEKTNESNMQKOLEHSNDALKAQ 1768
Qy 1085 -----QDSDROGNSLNDQ-----AGLLQKQKAILL-----GNQFIQYYQNDKEI 1126
Db 1769 DLKIQEBELIAHMLKQEQETIDKLRGIVSEKTDKLSNMQKOLENSNAKLQEKIQELKAN 1828
Qy 1127 EFETIINV-----EKVSLSFRVEPKLAKTLEDNGKTIRVLSDEMTSLIVNTTIEKTP 1178
Db 1829 EHQIITLKKDVNETQKVSWE-----QLKKQIKDQSLTSLKLEIENLNL-AQELHENLE 1882
Qy 1179 EMSAVPEVFT-KWVEQ-----YDPRTPLAATKFKVLKFKDQIPVDG 1219
Db 1883 EMKSVMERDNLNRVEETLKLERDQKESLQETKARDLEIQBLKTARMLSKHEKETVD-1941
Qy 1220 SGNISDKWLASIPLVIHQMLRLSPVVKTRIGLKTQEQOQQOQQOQQOQQOQKAVRKE 1279
Db 1942 --KLREX-----ISEKTIQISDIQK-----DLQSKDELOKKIQELQKELQLLRVK 1986
Qy 1280 BELETYPKDFEFTNLPLTKAHLRTLSNLVNDPNYKIEDLVKQNEAGDHQLAFSLRAN 1339
Db 1987 EDVNMSHKK--INEMEQKKQF-----EPNY-----LCKCEMDNFQITKTLH--2026
Qy 1340 NIKRLMNTPTTFADYNPFYYNEDWRSIDKYLANKGN--VSSHQQQAAGNQGSGLIQRL 1397
Db 2027 --ESLEEIRIVAKE-----RDELRIKESLSKMERDQFIATLREMIARDQ-----2069
Qy 1398 NKNIKPTFTPALIALKDRNNTLSNYSDKIIMKPKYLVERSIGVPMWSTGLDGYTQSEQ 1457
Db 2070 NHQVKPK-----RLUSDGQOHLMESLREKCSRIKE--LLKR-----YSEMDHY-----2112
Qy 1458 TKDGTSSSSQKQPDQFIOALGI---KNTEYHG-----KLGLSIRIFDPGNELAKIKDA 1509
Db 2113 -----ECLNRLSLDLEKEIEPHRMKKLVLSY-----VTKKEE 2148

1510 SNKKGEKLLKSYDLFPKNYLNEYKSPKIAKGTWNIHPDQKEYPNPNQKLPENYINLVL 1569
2149 -----QHECINKFEM--DFIDEVEKQKELLIK-----IQHLQDQCDVPSRELDRD---LKL 2193
1570 NQPKVKTL-----YNSDPF-----ITN-----LFVEPEGSDRSGTGLKQVIOK 1608
2194 NQMDLHIEILKDFSEBFPFIKTEFQOVLGNRKEMTOFLEEWLNTRFDIEKLNKGIQK 2253
1609 -----QVNNNVADMGSAVLTFTWYDKNI-ITNQPNVI---TANIADVFIKDVKEL-EDNT 1657
2254 ENDRICQVNN-----FNNRIIAIMNESTEFERSATISKWEQDLSLKEKNE 2302
1658 KUAPNITQWPNISGSKFYKPTVFFGNWENENSSMNSQAQPTPWKIRE 1709
2303 KLFKNYQTLKTSLSAQAQ---VNPT-----TQDNKNPHVTSRATQLTTEKIRE 2347

RESULT 13
ABG06301
ID ABG06301 standard; protein; 2017 AA.
XX ABG06301;
XX AC
XX DT 13-FEB-2002 (first entry)
XX DE
XX Novel human diagnostic protein #6292.
XX DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX XX
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX XX
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX XX
XX WPI: 2001-639362/73.
XX N-PSDB; AAS70488.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 36660; 103pp; English.
XX XX
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have application in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic

CC	amino acid sequences of the invention. Note: The sequence data for this									
CC	patent did not appear in the printed specification, but was obtained in									
CC	electronic format directly from WIPO at									
CC	ftp.wipo.int/pub/published_pct_sequences									
XX										
QY	Sequence 2017 AA;									
	Query Match	3.1%;	Score 299;	DB 4;	Length 2017;					
	Best Local Similarity	18.6%;	Pred. No. 4.2e-08;							
	Matches 368;	Conservative 354;	Mismatches 668;	Indels 584;	Gaps 95;					
QY	22	FGTVVGLASVKYGVNPTQGVISOLGLIDSVAFKPSIANFTSD-								YQSVKAL 73
DB	232	FGDIISQOETN-RLSNEVSLSEVGHWRHIA-QTSAQGTDSOSEICKLQNIKEL								289
QY	74	LANGTFDPKSEFTDFVSKFDLTN-NGRITVLEIPKPYQVVVISFSPEDDKERFLGFHL								132
DB	290	KQN-----RSQEIIDHGHMSVLQNAHQKLTETSRHRELSDY-----EERI-								333
QY	133	KEKLEDGNIASATKPIYLLPLDMPKALGOYSVIVDKVFNLLIHLPLSNFSAQSIKPLA								192
DB	334	-EELE--NLLQGGSGV--IETDLSKI-----YEMQKTIQVLEKVE--STKKMEOLE								380
QY	193	LTRSDFTAKLNQFN-----ODELVYLEKFFDLKAMIRLQ-----TA								234
DB	381	-DKIKDINKLSAENDRIILRREQELNV--EKQIMECE-NLKLECSKLQPSAVKQS								436
QY	235	DFSEFK-----GNLVDPP-VYSPIRNPQKEMASDLNQD-----QKTVRLYLIRTEFSP								282
DB	437	DTWTEKERILAQSGSVEEVRRLQALSDAENEIMRLSSLNQNSLAEADNLKLMRIEVL								496
QY	283	QAKTILKDYKYDFTFLSIDLK-----ASNGTSLPANENDLKQDLVDLLDVS DY								333
DB	497	KEKSLLS--QEKEELQMSLLKNNYEVIKSTATRDISLSDSELHDLRLNLE-----								545
QY	334	FGQSETITSNQVPPVPAESRLKDRVKEFKDQKPKRIEKFSLVEYDALSFYSOLOELV								393
DB	546	-----AKGELNQGIS-EKETLIAEIBELDRQONQATKHMILIKALQ								586
QY	394	SKPNSIKDVLNATLARNLFSLGKYNFLFDLASHLYYFLVSKAKIKOSSITKKLFIEL								453
DB	587	SKQNEGDSIISKLAQDLNDEKRVHQLDD-----KMDITKELDVQF								629
QY	454	PIKISLSGILGDQPNKITLPEKEVTFKLDNFRDVEIEKAFGLLYPGVNELEQARQAQ								513
DB	630	VLLIQSEVAL-----NDLHLTKQKLEKVENL-----VDQLNKSQ								664
QY	514	RASPEKSKGKLFESQKENSKAINNQ--EGLEEDN-----ITER-----LPE								558
DB	665	ESNVSIOKENLEKHEIRONEEELSIRNELMQSLNQDSNSNFKDTLLKEREAEVRNLKQ								724
QY	559	N-SPIYQOENAGLASDPKPYMKDVQNOYLLA-----KSQIBELIKAKDYTKLAKLLS								613
DB	725	NLSELEQLNENLKVAFDVK-----MENEKLVLACEDVRHQLEBCLAGNQLSLEK---								775
QY	614	NRHTYNISURLKEQLFDVNPRIPSSRDIEKAFVLDKTEKNQWQYSSASPVFQNKSL								673
DB	776	--NTIVETLMKEGTEAEELCAWAKRLLEA-----NKYEKTEIELSNA-RN----								819
QY	674	FGYYRILLGLDPKQTHIE-LVKLGOKAGLOF-----EGVENLPSPDF-----NLEDLK								719
DB	820	-----LNTSALQLEHEHLIKLNKKOMEIAELKNITBQMDTDHKEYKDVLSSELEQK								872
QY	720	NIRIKTFLPSQDNFKSLDLDFNNYDGEIKAPFGLPLFLPKELRRN-----SSN								770
DB	873	QL---TQLINKKEIFIEKLKERSKQLQELDK-----YQALRKNKEILRQTIEEKDR								921
QY	771	SGSQNSNFWPQEIISQPKDQ-----NLSNQDLA-----QFST-----K								806
DB	922	SLGSMKEENHLQEELERLREEQSRTPAPVADPKTLDVSTELASEVSQNLNTIKEHLEBEIK								981
QY	807	IWEKIIGDENEPQNNRLQYKLLKDLQESWINKTRDNLVYTLGDKLVKPKNNLEAKFR								866

DB	982	HHQKIIDQNO-----SKMQ--LIQSLQEQ--KKEMDE--FRYQHEQMNAHTHQLFLEKDE								1031
QY	867	QISNLOELLTAFYT-----SAALSNNWVYQDSGAKSTIPE								903
DB	1032	EIKSLQKTIQIKTQLHEERQDIQTNSDIFQSTKVQSLNIENGSEKHLDSKAETERLVK								1091
QY	904	EIAELDPKVK--EK--VGADVYQLKFHYAIGFDDNAGKFNQEV-----IR								944
DB	1092	GIKERELEIKLAKNEKISLTKQIDQLS-----KDEVGKLTQIIQOKOLEIQALHARIS								1144
QY	945	SSRT---IYLTSGSKLEADTIDQ-----LNOAVKNAPLGLQSPYLDTERFGVPQKL								995
DB	1145	STSHQTDVVYL-----QQOLQAYAMEREKPAVLNEKTR-----ENSHLTXEYHKMMDIV								1194
QY	996	ATSLA--VOHKQKEKTLPKKLNNDGYTLIHDKLKCPVPOISSSPKEDWPEGLKNQNGQS								1053
DB	1195	AAKEAALIKLODENKKLSTRFSSGQDMFRETQN--LSRIIR--EKDIDIDALSQKQCT								1250
QY	1054	-----QNVNVSTFGSIIESPYFTNFQEDADLD-----QDGQDDSRQG								1091
DB	1251	LLAVLQTSSTGNEAGGVNSNQFELLQERDKLQOVKKMBEKKQOVMTTVQNMQHESAQL								1310
QY	1092	NNSLDNQEAGLL-----KQKAILLGNQPIQYIYQNDKEI-----								1126
DB	1311	QEEHLQLOAQVLDVSDNNSKLQVDY-TGLIQSYEQNETKLNKFGQELAQVQHSIGQLCNT								1369
QY	1127	-----EFELIN-----VEKVSSELSFRVEFKLAKTLEDNGK								1156
DB	1370	KDLLGLKLDIISPOLSSASLLTPQSAECLRASKEVSESELQOELEBLRKSQEKDA								1429
QY	1157	TIRVLSD-----TMSLIVNTTITKTPMSAVPEVFDTKWVEQYDPRTPPLAAKTFFVLKFK								1212
DB	1430	TIRLOENHRLSDSIAATSELERKEHQDSEIKQLK--EKQDVQLKLLKEKOLLIKAK								1487
QY	1213	DQIPVDGSGNISDKW-----LASIPLVIHQOUMRL--SPVVKTIRELGLKT								1256
DB	1488	SDQLSSNENFTKNVENELLRQAVTNLKERILILEMDIGLKGENEKIVETYR--GKET								1545
QY	1257	EQOQOQOQOQOQOQOQKAVKREBELETYNPKDSFNILNPLTKAHLTSLNVLNNDPN--								1314
DB	1546	EYQALQDTNNKFSM--MLREKEFECHSMEKALAFQELLKEKEQGTGELQNLNAV								1600
QY	1315	YKIBDLKVIKNEAGDHQALFSLRANNIKRLANTPITFADYNPPFYFNDSIDKYLANK								1374
DB	1601	KSMQEKTVVQOERD-QVMLALKQ--KOMENTAL-----QNEVQRLDK-----								1641
QY	1375	GNVSSHQOQAGGQSGSLQRLNKNI--KPEFTTPALIALKDRNTVLSNYSKIMIK								1432
DB	1642	-EPRSNQ-----LERLNLHLESEDSYTRALAAEDRE---AKLRKKVTVLE								1685
QY	1433	PKYLVERSIGVPWSTGLDGYIGSOTKDGTSSSQOKGFODDFIQ-----ALGL								1481
DB	1686	EK-LVSSSNNAME-NASHQASVQVESLQELNVSKQR--DETALQLSVSOVQVQYALSL								1741
QY	1482	KNTEYHGKLGSLIRIFOPGNELAKIKDASNKKGEEKLLKSYDL--FKNVLNEYEKKSPI								1539
DB	1742	AN-----LQWVLEHF-----QOEEKAMYSAELEKQKQLIAEWKXQTKQI								1780
QY	1540	AKG-----WTNTHPQKEYPNPNOKLPENYLNVLNLPKPV								1575
DB	1781	LEGKVISLQECLEDEANAALDSASRLTEQLDKQEBQIEBKRLQNELRQEMLDDV--OKLM								1838
QY	1576	TLYNSD-----FITNLV-----EPGSDR-----GSGTKLKQVIOKVNNNYADWGS								1619
DB	1839	SLANSSEKGVKVLNRNLFIHFHTPKNQRHEVRLMGSILGVRENEEQLFHDDQGGVT								1898
QY	1620	AYLTFWY-----DKNIITNPQNPVITANIADVFIDKVKLEDNKTLIAPNT								1665
DB	1899	RWMTGCGGSGSKSPVNTPLRPNQOSVNSFSFELFVKFL-ETESHPSIPPPKLS								1951

Db 1349 P-----YKFLNKEKD---KFLSSYNIKSIDTIDFANDVLGYKILSEKYSKSLDS 1399

Qy 1367 IDKYLNNK 1374

Db 1400 IKKYINDK 1407

RESULT 15

AAE29345
ID AAE29345 standard; protein; 1639 AA.

XX AC AAE29345;

XX 27-JAN-2003 (first entry)

XX Plasmodium falciparum merozoite surface protein-1 (MSP-1).

XX Band 3 polypeptide; malarial infection; drug resistance; vaccine;
KW protozoacide; gene therapy.

XX Plasmodium falciparum.

OS WO200270542-A2.

XX 12-SEP-2002.

XX 01-MAR-2002; 2002WO-US006415.

XX 02-MAR-2001; 2001US-0272930P.

XX (SELI-) ST ELIZABETH'S MEDICAL CENT INC.

XX Chishti AH, Oh SS, Liu D, Goel V;

XX WPI: 2002-759814/82.

XX N-PSDB; AAD46980.

XX New isolated Band 3 polypeptide which selectively binds to merozoite
PT surface protein-1, useful for the prevention and treatment of malarial
PT infection.

XX Disclosure; Page 113-118; 163pp; English.

XX The invention relates to an isolated Band 3 polypeptide that comprises
CC any of 4 20 residue amino acid sequences, or their fragments that bind to
CC an MSP-1 polypeptide or a polypeptide with any of 8 291-1331 base pair
CC sequences. The methods and compositions of the present invention are
CC useful for the prevention and treatment of malarial infection. The
CC present invention, develops new and more improved methods based upon
CC inhibiting the particular interactions between the malarial parasite and
CC a cognate molecule present in the host and subsequently minimising
CC harmful side effects and drug resistance that may be due to non-specific
CC therapeutic approaches. The invention is useful in gene therapy. The
CC present invention also provides a vaccine for malaria. The present
CC sequence is Plasmodium falciparum merozoite surface protein-1 (MSP-1)

XX SQ Sequence 1639 AA;

Query Match 3.0%; Score 295; DB 5; Length 1639;

Best Local Similarity 19.5%; Pred. No. 5.3e-08;

Matches 282; Conservative 204; Mismatches 478; Indels 484; Gaps 68;

Qy 203 LNFQNNQDELWVYLEKFFDLEALKANIRLQADFSFEKGNLVDPFFVYSFIRPQKQKWA 262

Db 168 IDGVEEINELLYKLNFFDILLRAKLN-----DVCANDYQCQIPFNLKIRA 211

Qy 263 SDLANQOKTVRLYLRTFSPQAKTLKDYKDKETFLSSIDLKASNGTSLFANENDLQK 322

Db 212 NELDLVLLKGLV-----FGYRKPL-----DNIKDN 234

Qy 323 LDVLLDVSDFGGQGETITNSQVKPVPASERSLKDVRKFKDQKQPRIEK-----FSLY 378

Db 235 VG-----KMEDYIKNNKKTIENTINEL--IESKKTIDKNKNATKEEKKLYQAQYDLSY 288

Qy 379 -----EYDALSFYSQLELVSKPNSIKDLVNAI----- 406
Db 289 NKOLEBAHNLI SVLEKRIDTLKKNENIKELDKINEIKPPPPANSNGTPTNLLDKNKKIE 348
Qy 407 -----LARNIRFSLGKYNFLFDLASHLDYVFLVSKAKIKOSSITKKLFIELPIKIS 458
Db 349 EHEKEIKEIAKTIKFI--DSLFTD-PLELEYL--REKNKN-----IDISAKVE 393
Qy 459 LKSSILQDQEPNIKTIFKEVTFKLDNFRDVEIEKAFGLLYPGVNEELEQARK-----AQR 514
Db 394 TKESTEPNEYPN-----GVTYPLS-----YNDINNALNEINSFGDLINP 432
Qy 515 ASFEKEKSKGLKEFSQOKEENSKAINNOEGLBEDDNIETRLPENSPYIQOENAGLGAS 574
Db 433 PDYKEFSKNIYTD-----NERKFFNE-----IKEKIKIEKKIESDKKSYE----- 475
Qy 575 PDKPYMIKOVQORYYLAQSIOELIKAK-----DYTKLAKLSNRHUYNISRLKEQLP 629
Db 476 -DRSKSLNDITKEYEKL--LNEIYDSKFNNNIDLTNFEKMMGKRYSYKV-----EKL 525
Qy 630 DVNPRIPSSRDIEKAFVLDK-TEKNKYWQIYSSASPVFQNKWSLFGYVRYLLGL--DPK 686
Db 526 HHN-----TFASYENSCKNLEKLTAKALYMSDYSLRNIVVEKE--LKYKNLISKIENEI 578
Qy 687 QTIHELVLGOKAGLOPEGY---ENLPSPDNLE--DLKNIRIKTPIFSQK-DNFKLSLL 739
Db 579 ETLVENIKDBEQ--LFEKKITKDNKPEKILEVSDIVKVQVQVLLMKNKIDELKKTQL 636
Qy 740 DFNN-----YYDGBIKAPEFGPLPLPK-----ELRNSNSG 772
Db 637 ILKNVELKHNHVPNSYKQENKQEPYLI VLKKEIDKLVFMPKVESLINEEKNKITEG 696
Qy 773 GSQNSNPFWEOLISQF----- 789
Db 697 QSDNSEPSTEGETTGQATTKPGQOAGSALGDSVQAQAEQKQAPPPVPVPEAKAQVP 756
Qy 790 -----KQONLSNQDLAQFSTKIWE--KIIGDENEFQDNRLQYKLLKD 831
Db 757 TPPAPVNKNTEVSKLDYL-----EKLYEFLNTSVYCHYILVSHSTWNEKILAKYKITKE 812
Qy 832 LQESWINKTRDNLWYTLGDKLVKPKKNLEAKFQI-----SNLOELLTAPYTSAAASNN 887
Db 813 -EESKLSK-DPL-----DLLEFNQNNIPVMYSMFDLSNLSQLFMEIYEMVNCNL 863
Qy 888 WNYQDSGAKSTIIFERIAELDPKVEKGVADVQLK----- 924
Db 864 YKLKDNKIKN--LLEAKKVSTSVKTLSSSMQPLSLTPQDKPEVSANDTSHSTNLMN 921
Qy 925 ---FHVAIGFDDNAGKFNQEVIRSSRTIYLTSGKS-----KLEADTIDQL 968
Db 922 SLKLFENILSGKKNYIQLIGQKSENFEYKILKSDTFYNESFTFVFSKADDSINSL 981
Qy 969 NQAVKNAPLGLQSFYLDTERFGVFOKATSLAVQHKOK-EKTLPKLNNDGYTLIHDKLK 1027
Db 982 NDESKRKKL-----EEDINKLKTQLSFDLYNKYKLERLFDKKTGVGYKM---QIK 1033
Qy 1028 KPVIPOISSSPKDFEGLKNQSGQNV--NVSTFGSIIESPVSTFQEDADLDQDQ 1085
Db 1034 KLTL-----LKEQLESKLSLNNPCHVLQNFVSF-----FNKKKEAEI----- 1071
Qy 1086 DDSRGNNSLDNOBAGLLKQKLAILLGNQFIQYQOND---KEIFEIINVEK--VSELS 1140
Db 1072 ---AETENLENTKI-LLKHYKGL-----VKYNGESSPLKLTSEESIQTEDNYASLEN 1121
Qy 1141 FRVEPKLAKTLEDN---GKTIRVLSDETSLI-----VNTT 1173
Db 1122 FAVLSKLSGKLKDLNLEKKKLSYLSGLHLLIABLKEVKNKONYTGNSPSENNTDVNA 1181
Qy 1174 TK-----TPMSAVPEVF-----DTKWVEQYDPRTP---LAAKTKFVLKPKDQ----- 1214
Db 1182 LESYKFKLPEGTDVATVVSSEGSOT--LEQSQPKPASTHVGAESNTITTSQNVDEVD 1239

```
QY 1215 ---IPVDGSGNISDKWLASI-----PLVI-----HQMLRLSPVVKTIIRGLGK 1255
DB 1240 VIIVPIFGSESEYDDLGQVVTGEAVTPSIDNLSKIENEYELVYKPLAGVYRSL--- 1296
QY 1256 TEQQOQQOQQOQQOQKKAVERKEELET-YNPKDEF-NIL-NPLTKAHLRLTSLNVNND 1312
DB 1297 -----KKQLENNVWFNVNVKDIILNSRFKRENFKNVLESDLIPYKDLTSSNVVVD 1348
QY 1313 PNYKIEDLVKIRNEAGDQHLAFSLRANNIKRLMNTPIITFADYNPPFY-----YNBDWRS 1366
DB 1349 P-----YKFLNKEKRD---KFLSSYNYIKDSIDTDINPANDVLGVYKILSEKYKSDLDS 1399
QY 1367 IDKYLANK 1374
DB 1400 IKKYINDK 1407

RESULT 16
AAP50777
ID AAP50777 standard; protein; 1654 AA.
XX
AC AAP50777;
XX
DT 25-MAR-2003 (revised)
DT 30-SEP-1991 (first entry)
XX
XX Sequence of the P195 protein of Plasmodium falciparum.
XX
XX Malaria vaccine; epitope; antigen; immunogen.
XX
OS Plasmodium falciparum.
XX
PN EP154454-A.
XX
PD 11-SEP-1985.
XX
PF 21-FEB-1985; 85EP-00301173.
XX
PR 22-FEB-1984; 84GB-00004692.
PR 26-SEP-1984; 84GB-00024340.
PR 21-FEB-1985; 85GB-00004429.
XX
XX (WELL ) WELLCOME FOUND LTD.
XX
XX Holder A, Sandhu J, Odink K, Lockyer M, Riverosmor V;
XX
XX WPI; 1985-224845/37.
XX
XX N-PSDB; AAN50530.
XX
XX Cloned DNA sequence encoding plasmodium falciparum protein - useful for
XX expressing the protein for use in vaccines against malaria.
XX
XX Claim 6; Fig 1; 51pp; English.
XX
XX The sequence encoding the P195 protein of Plasmodium falciparum
XX (AAN50530) and a peptide comprising at least one of its epitopes (see
XX AAP50777) are claimed. Also claimed is a vaccine for inducing immunity to
XX malaria comprising the novel peptide or P195 or a peptide comprising at
XX least one epitope when derived from the new DNA sequence, together with a
XX carrier. (Updated on 25-MAR-2003 to correct PA field.) (Updated on 25-MAR
XX -2003 to correct PI field.)
XX
XX Sequence 1654 AA;
XX
XX Query Match 3.0%; Score 295; DB 1; Length 1654;
XX Best Local Similarity 19.2%; Pred. No. 5.4e-08;
XX Matches 324; Conservative 223; Mismatches 547; Indels 590; Gaps 78;
QY 203 LNOFNQDELWVLYLEKFFDLEALKANIRLQTADFSFEKGNLVDPFVYSFIRNPQOKEWA 262
DB 168 IDGYEINELLYKLNIFYDLRLAKLN-----DVCANDYCQIPFNLKIRA 211
QY 263 SLDNQDQKTVRLYLRTEFSPQAKTILKDYKYKDETFLSSIDLKASNGTSLPANENDLKDQ 322
```

```
DB 212 NELDVLKLV-----FGYRKPL-----DNIKON 234
QY 323 LDVLDLVDYFGGOSSETITTSQVKVPASBRSLKDRVKFKDQOQPRIEK-----FSLY 378
DB 235 VG-----KMEDIYKKNKTNIENEL--IEESKKTIDKNKNATKEBEKKLYQAOYDLSIY 288
QY 379 -----EYDALSFYQLOELVSKPNSIKOLVNAI----- 406
DB 289 NKQLEBAHNLSVLBEKRIDTLKKNENIKELLKINEIKNPPANSGNTPTNLLDKNKKIE 348
QY 407 -----LARNLRSLGKYNEFDLASHLDYFLVSKAKIKQSSITKKLFIELPIKS 458
DB 349 EHEKEIKEIAKTIKFNI-----DSLFTD-PLELEYL--REKNKN-----IDISAKVE 393
QY 459 LKSSILGDOEPNIKTILFEKEVTFKLDNPRDVBIEKAFGLLYPGVNEELQARK---AOR 514
DB 394 TKESTEPNEYFN-----GVITYPLS-----YNDINNALNELSGDLINP 432
QY 515 ASFEKEKSKGLKEFSQOQKEBNSKAINNOEGEEEDDNIITRLPENSPIOYOENAGLGAS 574
DB 433 FDYTKEPSKNIYTD-----NERKKFINE---IKEKIKIEKKIESDKSYE----- 475
QY 575 PDKPYMIKDQVQRYVLAKSOIQELIKAK-----DYTKLAKLLSNRHTYNISLRKEQLF 629
DB 476 -DRSKSLNDITKEYEKL-----LNEIYDSKFNNDILTNFEKMMGKRYSYKV-----EKL 525
QY 630 DYNPRIPSSRDIEKAFVLDK-TEKNKYQWQIYSSASPVFQNKWSLFGYVYRLLGL--DPK 686
DB 526 HFN---TPASYENSKHNLEKLTALKYMEDYSLRNIVVEKE---LKYYKNLISKIENEI 578
QY 687 QTIHELVLKGQAGLOFEGY-----ENLPSDFNLE--DLKNIRIKTLPFSQK-DNFKLSLL 739
DB 579 ETLVENIKKDEBQ--LFEKKITKDNENKPKDEKLEVSIVKVQVQVLLMKNKIDELKKTQL 636
QY 740 DFNN-----YDGEIKAPEFGLPLFLPK-----ELRRNSNSG 772
DB 637 ILKNVELKNIHVNPNSYKQENKQEPYILVLKKEIDKLKVPKPKVESLINEKKNIKTQG 696
QY 773 GSQNSNPWEQIISIQF----- 789
DB 697 QSDNSEPSTEGEITGQATTKPGQAGSALGSDVQAQAEQKQAOQPPVPVPEAKAQP 756
QY 790 -----KDQNLNQD---QLAQF--STKIWEK-IIGDENEFQNNRLOYKLLKDLQES 835
DB 757 TPPAPVNNKNTENVSKLDYLEKLYQFLNTSYICHKYILVSHSTMNEKILKQYKITKE-EES 815
QY 836 WIN-----KTRDNLVWTY-LGDKLVKPKNNLEAKFROISNLQELLTAPYTSALS 885
DB 816 KLSGCDPLDLLFNQNNIPVMYSMFDLSLNIYHNLVWYVEKEIGYVFIILLMEIYEMVC 875
QY 886 NNWNYQDSGAKSTIIFEEIAELDPKVKQKGVADYVYQLK----- 924
DB 876 NLYKLDNDKIKN--LLEEAKKVSTSVKTLSSSQPLSLTPQDKPEVSANDTSHSTNL 933
QY 925 -----FYAIGFDDNAGKQOEVRSSRTIYLTSGKS-----KLEADTID 966
DB 934 NNSLKLFIENILSLGKNKNIYQELIQKSENSEYKILKDSDTFFYNESFTNFVKSKADDIN 993
QY 967 QLNQAVKNAPLGLQSFYLDTERFVGFKLATSIAVQHKQK-EKTLPKKLNNNGYTLIHDK 1025
DB 994 SLNDESKRKKL-----EEDINKLKTQLSFDLYNKYKLERLPDKKTVGVKYM---Q 1045
QY 1026 LKKPVIPOISSPEKDFEGKLNQNGSQNV--NVSTFGSIIIESPYFSFNFQEDADLDQD 1083
DB 1046 IKKLTJL-----LKEQLESKLSLNNPKHVLQNFVSF-----ENKKEAEI--- 1085
QY 1084 QODDSRQGNNSLDNQEAGLLKQKLAILLGNQFIQYQQND---KETEFILINVEK--VSE 1138
DB 1086 -----AETENTLUENTKI-LLKHVYKGL-----VKYNGESSPLKTUSEESIOTEDNYASL 1133
QY 1139 LSFVREFFKLAKTLEDN---GKTIRVLSDETMSLI-----VN 1171
```

Db 1134 ENFKVLSKLEGLKDNLNLEKKKLSYLSRGLHLHIAELKEVINKNKNYTGNSPVSNNTDVN 1193

Qy 1172 TTIEK----TPEMSAVEPV-----DTKWVEQVDPRTPLAAKTKFVLKFKDQIPVDGSGN 1222

Db 1194 NALESYKKFLPEGTDVATVVSSESGDT--LEQSQPKKPASTHVG-----AESNTITTSQN 1246

Qy 1223 ISDKWLASIPLVI-----HQQMLRLSPVVKT 1248

Db 1247 VDDEVDDVILVIFGESEEDYDGLGVVTEAVTTSVIDNLSKIENEYEVLYLKPLAGV 1306

Qy 1249 IRELGLKTEQOQQOQQOQQOQKQKAVRKEELET-YNPKDEF-NIL-NPLTKAHLTL 1305

Db 1307 YRSL-----KKQLENNVMTFVNVVKDILNSRFNKEFNKNVLESDLIPYKDLTS 1355

Qy 1306 SNLVNPNYKIEDLVKIKNAGHQHAFSLRANNIKRLMNTPTITPADYNPFY----- 1359

Db 1356 SNYVVKDP-----YKFLNKEKRD---KFLSSYNYIKDSIDTIDNFANDVLGYKILSEK 1406

Qy 1360 YNEDWRSIDKYLNNKGNVSSHQQQAAGNQGSLIQRLNKNIKPETFPTPALIALKDRNN- 1418

Db 1407 YKSLDLSIKKYINDK-----QGEN-----EKYLPFL-----NNI 1435

Qy 1419 -TNLSNYSDKI-----IMIKPK---YLVERSIGVPWSTGLDGYIGSEQTKDGTSSSSQK 1470

Db 1436 ETLKYTVNDKIDLFIHLEAKVLYTEKS-----NVEVK----- 1470

Qy 1471 FDQDFIQALGNTEYHGLGLSIRIFDPGNELAKIKDASNKGEKLLKSY----- 1522

Db 1471 -----IKELNLYKTQDKLADPKNNNFVGIADLSYDYNHNNLLTKFLSTGMVFE 1520

Qy 1523 DLFKNYL-NEVEKSPKIAKWTN-----IHPDQKE-----YPNPNQK 1559

Db 1521 NLLKSVLSNLDWKLARYVGHFTTPMRKKTWIOSSGCFRHLDERECKLLNLYKQSGSK 1580

Qy 1560 LPEN 1563

Db 1581 CVEN 1584

RESULT 17

AAR05591

ID AAR05591 standard; protein; 896 AA.

XX AAR05591;

AC AAR05591;

XX

XX

DT 25-MAR-2003 (revised)

DT 07-AUG-1990 (first entry)

DE

XX The 36 kD antigen of Mycoplasma hyopneumonia.

XX Recombinant Mycoplasma hyopneumoniae antigens; mycoplasma pneumonia;

KW vaccine; antibody; da.

XX

OS Synthetic.

XX

FH Location/Qualifiers

FT 78..78

FT /label= Coded by TGA nonsense codon.

FT 292..292

FT /label= Coded by TGA nonsense codon.

FT 298..298

FT /label= Coded by TAA nonsense codon.

FT 366..366

FT /label= Coded by TAA nonsense codon.

FT 370..370

FT /label= Coded by TAA nonsense codon.

FT 377..377

FT /label= Coded by TGA nonsense codon.

FT 414..414

FT /label= Coded by TAG nonsense codon.

FT 420..420

FT /label= Coded by TAA nonsense codon.

FT 527..528

FT Region

FT 534..534

FT /label= Undefined region of 88 units long.

FT 539..539

FT /label= Coded by TAA nonsense codon.

FT 549..549

FT /label= Coded by TAA nonsense codon.

FT 559..559

FT /label= Coded by TAA nonsense codon.

FT 664..664

FT /label= Coded by TAA nonsense codon.

FT 676..676

FT /label= Coded by TGA nonsense codon.

FT 677..677

FT /label= Coded by TGA nonsense codon.

FT 725..726

FT /label= Undefined region of 50 units long.

FT 729..729

FT /label= Coded by TGA nonsense codon.

FT 734..734

FT /label= Coded by TAA nonsense codon.

FT 743..743

FT /label= Coded by TGA nonsense codon.

FT 744..744

FT /label= Coded by TAA nonsense codon.

FT 756..756

FT /label= Coded by TAA nonsense codon.

FT 788..788

FT /label= Coded by TAA nonsense codon

FT 790..790

FT /label= Coded by TGA nonsense codon.

XX

XX EP359919-A.

XX

XX 28-MAR-1990.

XX

XX 28-JUN-1989; 89EP-00111748.

XX

XX 29-JUN-1988; 88US-00213248.

PR 07-APR-1989; 89US-00334586.

PR 21-APR-1989; 89US-00341968.

XX

XX (MLTE-) ML TECHN VENTURES L.

XX

XX Faulds DH, Brooks E, Andrews WH, Lory C;

XX

XX WPI; 1990-092610/13.

DR N-PSDB; AAQ03634.

XX

XX Recombinant Mycoplasma hyopneumoniae antigens - used in vaccines against

FT mycoplasma pneumonia and in assays for detection or determ. of

FT antibody.

XX

XX Disclosure; Fig 34; 54pp; English.

XX

XX An expression vehicle containing the sequence is capable of eliciting an

CC antibody which recognises an epitope of an M. hyo antigen. The

CC recombinant proteins can be used in a vaccine for protection against

CC Mycoplasma pneumonia, partic. in swine, br in an assay for detection of

CC antibodies to M.hyo. All X's in the sequence that are not coded by

CC nonsense codons are blanks in the specification. See also AAQ03629-

CC Q03634, AAQ04959-Q04963 and AAQ02474. (Updated on 25-MAR-2003 to correct

CC PR field.) (Updated on 25-MAR-2003 to correct PA field.)

XX

SQ Sequence 896 AA;

Query Match 3.0%; Score 291.5; DB 2; Length 896;

Best Local Similarity 21.8%; Pred. No. 3.4e-08;

Matches 131; Conservative 109; Mismatches 240; Indels 121; Gaps 22;

Qy 1 MNKKSTLLATATAAIGSTVFGTVGLSKVKYRGVNPOTGVISQLGLDSVAFKPSIA 60

Db 1 MSKSKTFKIGLTAGIVGLVGLTVGLSLAKYRSBSPRKIANDFAAKAVSTLAFSPYAF 60

Db 1001 -----DLVSNFV-VLNYGVNEYFNLDLKH 1027
Qy 429 LDYVFLVSKAKIOSSITKKLFIELPIKISLKSILGDQEPNITKLFKEVETFKLNFDRD 488
Db 1028 RIYSF-----KKLIYEPNQNDKFSFLANQKIT-----NTEKT----- 1061
Qy 489 VEIEKAFGLLYPGVNEBLEQAKORASPEKESKGLK-----EFSQO-----K 533
Db 1062 -----ASVNEQLTIDTKLVKRP-DSNLKINLKLQINDPNDLFQDSILETFH 1109
Qy 534 EENSKAINNQEG-LEEDNITERLPENSPYOQENAG-LGASDPKPYMIKOVQORYYL 591
Db 1110 DEKDKTHNVIGKINVDAN-----NNTLEFSVENTNTFKIQNHXYIVDNIN-----YA 1159
Qy 592 AKSOIQELIKAKYTKLAKLSNRHT-----YNIS-----LRLKEQLF-----D 630
Db 1160 TENKIOP-----ANALSNNSKNISYDASSNPSKILSFTNELFVNINNVNPNNTN 1209
Qy 631 VNP-----RIPSRDIEKAFV-----LDTEKNKYQIYSSASPVF----- 667
Db 1210 LNPTASIDVELKSSQLLKQOYLRLALYIDNNHQ-KIWSDYASVNVNVDLAHINLNFHLTP 1268
Qy 668 QNKSLFGYVYLLGLDQPKQITHEL--VKLGKAGLOFEGYENLPSPDNLEDKNIIRIKT 725
Db 1269 NRKYNFAGIYF-----NDQONNDETQKKIFVKESKTYDPSTLASTTQLLEFFKTHV-- 1323
Qy 726 PLFSQONFKLSLLDFNYYDGEIKAPFGGLPLPKELRRNSNGSGSONSPWQOEI 785
Db 1324 -----DEDOFN--YDFNINDDQVLEPGMLVELCF----- 1351
Qy 786 ISQFKDQNLNQDQ-LAQFSTKIWEKILIGDENEDQNNRLQYKLLKOLQESWINKTRDNL 844
Db 1352 -----ENLVDKNOPVLATQTTLVKK-----DEHSFSAGVI-----TNLLADHT 1390
Qy 845 YWYLGDKLVKPK-----NNLEAKFQIQLNQLLELTAFYTSAAALSNNMNYQDSG 895
Db 1391 Y-RLTSVSLKQKQPLANVINNNNNNEILLNIEDNQTIHTL----- 1432
Qy 896 AKSTIIFBEIAEL-DPKVKEKGVADVYQLKHYAIGFDNAGKQEVIRSSRTIYLT 954
Db 1433 --SKSIVNEISNLADAYPKNNNG--IYDVKEN--INIIKNNKLNKYKVVVFEDNEHQL 1486
Qy 955 SKKSLEADTIDQLNAQVKNAPLQLOF-----YLDTERFGVFOKLAT 997
Db 1487 ISTNDLLVNLKDOTNL-----ISLQNFSPNLKPNHLYRLKVVYGEQNF----- 1533
Qy 998 SLAVQHKQKEKTLPKKLANDGYTLTHDKLKPVIPOISSPEKDWFEGL-----N 1048
Db 1534 --AINEQKNILALNPSLVNSFSSTTPAKIK-----VSKNAIDVWQOQALLIKLILDDSDN 1585
Qy 1049 Q--NGSQSNVNVSTFG--SIIESPYFSTNFQEDADLDQGDQDSDSQGNSLQNBAGLLK 1104
Db 1586 QLHSGDEINIRYKGTQNIISTPATIS-----DQDKRYTKCMATNLI-----AGLDY 1633
Qy 1105 QKLAILLGNQFTQ-----YQOQNDKEIFEI-I 1131
Db 1634 EIVSVTIKQKTKNSPIIFELPSPIGYTLAPVKITSMDPVYHENKNTADKLNL 1693
Qy 1132 NVEKVSLSFRVEFKLAKTLEONGTIRVLSDETMSLIVNTTIKTPEMSAVPEVFTKW 1191
Db 1694 RIENIGASLNFNDIKFIFKKNQDQKQISFIHKVT-----SANDANYEWEF 1738
Qy 1192 VEQYDPRTPLAARTKFLVLFKD--QIPVDGSGNISDKW-----LASIPLVHHQ 1238
Db 1739 KOLLRRNREYTLERVVY-LANKDFNQSNGVSESYI-DLWYIENLNKTFKLLPTKPLGI--- 1793
Qy 1239 MLRLSPVVTIRELGLKTEQO-----OQOQOQOQOQOKKAVRKEBELTYNPK 1288
Db 1794 ---IGAPIKEISDNGAKVLKFAINDFDVLKENQTFKNIQPNENNANLNEISEHECK 1850
Qy 1289 DEFNINLPTKHAHRLTSLNVLNNDPN--YKIBDLKVINE---AGDHOLAFS---LRANN 1340
Db 1851 VE-----IIDGQKFPFVLANLNKVNKEYKVKYKIFDENQDVNGVYKINFKNKYKEPN 1904

Qy 1341 I--KRLMNTPTITFADYNPFY--YNEDWESID-----KYL 1371
Db 1905 VVYDASVNTTQTVFTTKFAIASFNNLTDVAVANKQNISINLDSRVETIQGVHFKAKYI 1964
Qy 1372 -----NNKGNVSSHQOAGGNGSGSLIQL----- 1397
Db 1965 SINDRVVVTNTTIPAPTINGNKNKNIALNPEL-----NONQLISNRLYTFALYYSKDVND 2020
Qy 1398 -----NKNIKPTFP--TPALIALK-----DRNNTNLS-----NYSDKIMIKPKYLV 1437
Db 2021 ENHANMVIVKNNVNPOTISTKPSSTYVDLKAQNADENKITLSLLLSHNDQIFEDKNNHLK 2080
Qy 1438 ERSIGVPMSTGLDGYIGSQTGDTSSSQSGKQDFDQALGLK-NTEYHGK----- 1489
Db 2081 IAKISIDELDAHQIINS--TTHDYDYLTEKENNEWLLKTQLINLKNPTKYRVKVKWFTSK 2139
Qy 1490 -----LGLSIRIFDPG-----NELAKIKDASNKK 1513
Db 2140 PSDTIYGVNKNIVYPINHNNTNIDLKTLEKSTLANSVKPTKTNFNKNQESIKLMVGFNKT 2199
Qy 1514 G-----BEKLLKSYDLFKNYLNEYEKSPKIAKGWNTNIHPDOKEYPNPNQKLPENYLVNLVLN 1570
Db 2200 GSSLENKYAKL--VYKDNNNQIIESDVLNLSDTTTPKFNKEFSFN-----NOTSLVAN 2252
Qy 1571 QPW---KVTLYNSDF-----ITNLF-----VEPEGSDRGSGTKLKQVIQOVNN 1612
Db 2253 RQFEVKLIISDTADFQNTANTLDTNFKNDKDAKFSIEP-----TPISVNNVIOGSNT 2306
Qy 1613 NYADWGSAYLTFWY--DKNIITN-----ITANIADVFIKOVKELEDN 1656
Db 2307 -----YDEIHLSFDYDDQDHLVDNDQITITIRYKKEQAWTSKAGEVSVKNHKINVTLD 2362
Qy 1634 -----QPNV-----ITANIADVFIKOVKELEDN 1656
Db 2363 LTPNTTYEIGAISTTHVSPNKTSPIOYNETTKQLEKLDLTTKVAHNFIKDPK-LASN 2421
Qy 1657 TKLI-----APNITQWPNISGSKEKFKYKPTVFF 1685
Db 2422 SKLVGVYECADENKVFVGLHKEGHELEGYELIYAFVDTKTN-----KPAIEI 2472
Qy 1686 GNWENENSSNNSOATPTWEKIREGFALQALKSSFDQKTRFVLTTNAPLPLWKYGLGF 1745
Db 2473 KSSNNINSN-----QVYEFELLKSO-----ITPNHQVSELSICAEN 2511
Qy 1746 QNGENFTQDWRVLFQNDNQIAALRV-----QEQDRP-----EKSEDKDKQKWKFKVVI 1797
Db 2512 KGANNEIKKO-----LNENHLSLALGLAPNKMVDQKPTITITIKVDNQSVSAHVKIIND 2566
Qy 1798 PEEMFNSG-----NIRFVGWMOIQGPNTLWLPVINSVIYD-----FYRGTDGSDNVA 1845
Db 2567 PDGLINSDTSCSKELLNKKYALFNQTNANL-----INNITIKYDOONKQFY-----CEPDID 2618
Qy 1846 NLNV-APQVKTIAFTNNAFNNVKEFN 1872
Db 2619 NLTLNQDYGIFEISFANKPIHAAPAKIN 2646

RESULT 19

ADSA43650
ID ADS43650 standard; protein; 1679 AA.

XX ADS43650;

AC ADS43650;

DT 02-DEC-2004 (first entry)

XX Bacterial polypeptide #22080.

XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;

nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
bacterial polypeptide.

Bacteria.
US2003233675-A1.
18-DEC-2003.
20-FEB-2003; 2003US-00369493.
21-FEB-2002; 2002US-0360039P.
(CAOY/) CAO Y.
(HINK/) HINKLE G J.
(SLAT/) SLATER S C.
(CHEN/) CHEN X.
(GOLD/) GOLDMAN B S.
Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
WPI; 2004-061375/06.
New recombinant DNA construct comprising a promoter positioned to provide
for expression of a polynucleotide encoding a polypeptide from a
microbial source, useful for producing plants with improved properties.
Claim 1; SEQ ID NO 22080; 122pp; English.
The invention relates to a recombinant DNA construct comprising a
promoter functional in a plant cell, where the promoter is positioned to
provide for expression of a polynucleotide encoding a polypeptide from a
microbial source. The invention also relates to a transformed plant
comprising the recombinant DNA construct and a method of producing a
transformed plant having an improved property. The plant is a crop plant
such as maize or soybean. The method of producing a transformed plant
having an improved property comprises transforming a plant with the
recombinant DNA construct and growing the transformed plant, where the
polynucleotide or polypeptide is useful for improving plant properties.
The recombinant DNA construct is useful for producing plants with
improved plant properties, e.g. improved cold, heat or drought tolerance,
increased resistance to plant disease, better growth rate by modification
of the cell cycle pathway with plant growth regulators, increased rate of
homologous recombination, modified seed oil or protein yield and/or
content, improved yield by modification of carbohydrate, nitrogen or
phosphorus use and/or uptake, by modification of photosynthesis or by
providing improved plant growth and development under at least one stress
condition, improved lignin production or improved galactomannan
production. This sequence represents a bacterial polypeptide used in the
scope of the invention. Note: The sequence data for this patent did not
form part of the printed specification but was obtained in electronic
format from USPTO at seqdata.uspto.gov/sequence.html.
Sequence 1679 AA;
Query Match 3.0%; Score 290.5; DB 8; Length 1679;
Best Local Similarity 19.7%; Pred. No. 1e-07;
Matches 383; Conservative 312; Mismatches 684; Indels 567; Gaps 97;
QY 63 TSDYQSVKALLNGKTFDPKSSFTDFVSKFDFLTNGRTVLEIPKYQVVISERSPE-D 121
DB 69 SSSQKNKAEELNGK-DQLNEERSYRREIDAL-----KKQLHVSHEAMREYN 116
QY 122 DKERFRLGPHLKEKLELDGN-----IAQSATKFIYLLPLDMPKALGOVSY-- 166
DB 117 DEKRVKEEYDQSRDQNDLNDLKENKLLRRKLMENILQCKNSALSLQKYDT 176
QY 167 -IVDKNP-----NNLIHPLSNFSAQSIKPLALTRSS---DFTAKLQNF-NNODELMVY-- 215
DB 177 SVQEKELMLQSKKLIBEKLSSFSKTLTB-EVTKSSHVENLEBKLYMQMSVESVFTYNK 235
QY 216 -----LEKFFDLKALKANIRLOTADFSFKGNLVDPPVYSFIRNPQKQKWA 262

DB 236 FLLNQKQLSQSVBEKVEMKNLKDPTASVEKAFFSKB-----MTLQKNM 279
QY 263 SDLNQDQKTVRLYLRTFEPSPQAKTILKQYKXDETFPLSSIDLKASNGTSLFANENDLKQ 322
DB 280 NDLLRSQLT---SLEKDCSLRAIEKNDNSCRNPEHTVDVIDELIDTKLREKSKNCSQ-R 335
QY 323 LDVDDLVDSDYFGQSETITTSNQKVPVPASERSLKORVKFKDQKQKPKRIEKFSL---VE 379
DB 336 LQNIWMDCTK---EBEATMTTSVSPVTG--KLFSDIKVLKRQLIKERNQKFLQNLQE 389
QY 380 YDALSFPYQLOELVSKPNSIKDLVNATLARNLRFSLGKYNFLFDDLASHLDYVFLVSKAK 439
DB 390 DFLELEHKTPELISFKERTKSLH-----ELKRTELELTVSLTK 430
QY 440 IKQS-SITKKFLFELPIKISLSKSSILGDOBPNIKTLFEKEVTFKLDNFRDVEIEKAFGLL 498
DB 431 RKQEREIT-----SLRQKING-CEANIHSLVKQ---RLDLARQVKL-----LL 469
QY 499 YPGVNEELEQARKAQRASPEKEKSKGLKBFQSQEENSKAINNOEGLEEDDNITERLPE 558
DB 470 ---LNTSAIQETASPLSQDELISLRKILESSNIVNENDSOAI-----ITERLVE 515
QY 559 NSPI-QYQENAGL-----GASPDKPYMIKDVONQRYVYLAKSIOEL--IK 601
DB 516 FSNVNELOEKNVELLNCIRILADKLENVGKQDK--TLQKVENQTIKEAKDAIIELENIN 573
QY 602 AKDYTKLAKLSNRHTYNIISLRKQLFDVN-----PRIPSSRDIEKAKFVLDRTEKN 654
DB 574 AKWETRIILLRERDSYKLLASTEENKANTSVTSMEAREKKIRELE-AELSSTKVEN- 631
QY 655 KYWQYSSASPVFQNKWSLFGYRYLLGDPKQTHLVLKQKAGLQEGEYENLPSPDN 714
DB 632 -----SAIQN-----LRKELLYKKSQCKKT-----T 655
QY 715 LEDLNIRIKTFLFSQKNFLSLDDFNYYDGEIKAPEFGLPLFLPKELRNSNSGGS 774
DB 656 LEDFENFK---GLAKEKERMLEEAID---HLKAELEKQKSWPVSYIHVEKRASTELSQS 709
QY 775 QNSNPWEQEIISQFQDNLS---NQDQLAQSTTKIWEKIIGDNEFDQNNRLQYKLLKD 831
DB 710 RIKIKSLEVE-LSLKKKETASPIPTKESLTR-----DFEQCKEKKELQMR 754
QY 832 LQESWINTKRDNLWYTLGDKLVKPK---NNLEAKFQIISNLQELLTAFYTSAAALNN- 887
DB 755 LKESBISHNENKMFDSKSGEQYKAKIKELENNLE---RLRSDLSQKIQIESIRSCKSQ 811
QY 888 --W--NYVODSAGKSTIIFEEIETAEIDPKVKKGADVVQL-----KPHYAIGPDNAG 936
DB 812 LKWAQNTIDDTMKWKSLLTELUNKETTI-BKLSSEIENLDKELRKTFFQYKF-LDQNSD 869
QY 937 KFNQE-VIRSSSRTIYK-TSGKSKLEADTIDQLNQAVKNAPLQSLQSYFLDTEREGVFQK 994
DB 870 ASTLEPTLRKELEQIQVQLKDANSQIQ--YEEIISSENALIELKNELAKTK-----EN 922
QY 995 LATSIAVQHKQK---EKTLPKLNNDG-YTLIHDKLKPKVPIQIISSEPKOWFEGLKNQ 1050
DB 923 YDAKIELEKKEKWAREEDLSRLRGLGETRALQPKLKEGALHFVQOS-----EKLNRV 976
QY 1051 GOSQNV--NVSTFGSIIIE-----SPYFSTNFQSDAD-----LDQDQDSDSRGNNNS 1094
DB 977 ERIQMIKIEKQSTIVQLCKKKEWSQYOST-MKENKOLSELVIRLEKDAAD----- 1027
QY 1095 LDNQBAGLLKQKAILLGNQFIQYQQN--DKIEIFE---IINVEKVSSELSFRVE----- 1144
DB 1028 ---CQAEUTKTKSSLYSAQDLDLKHKKWMEBKADYERELISNIEQTE--SLRVENSVL 1082
QY 1145 FKLAKTLEDNG-----KTRIVLS-----DETMSLIVN-- 1171
DB 1083 EKVDVDTAANGDKHLLVSLFSLNLRHNSLETKLTTCKRELAFVKQKNSLEKTINDL 1142
QY 1172 ----TTIEKTPMSAVPEVFD-----TKWVEQYDPTPLAAKTKFVLKFDQIPVDGSGN 1222

Dd		1143	QRTOTLSEKYQCSAV--IIDEFKDITKEVTQN-----ILKENNAILQKSLKN	1189
Qy		1223	ISDKWLASIPLVIIHOQM-----LRLS-PVVKTIRELGL-----KTEOQQOOQQQQ	1266
Dd		1190	VTEK-----NREIYQLNDRQBETISRLORDLIQTKEQVSINSNKILVYESMEQCKORYQ	1244
Qy		1267	QQQPQPKAVRKBEELTYNPKBDEFINLPITKAHRLTLSNLVNPNPNYKIEDLKVIKNE	1326
Dd		1245	DLSQQQXDAQKKD-----IBKLTN-----EISDLK-----	1269
Qy		1327	AGDHOLAFSIRANNIKRLMNTPTFADYNPFYYNWEDWRSIDKYLNKNKG-----VSSHQQQ	1383
Dd		1270	---GKLSAENAN-----AD-----LENKFNRLLKQAHEKL	1297
Qy		1384	AAGNQSGGLIORLKNNIKPEFTFPALIALAKDRNNLTNSYDSKIIMIKPKYLVERSIGV	1443
Dd		1298	DASKKOAAALTNELNE-----LKAIDKLEODLHPENAKVI-----	1333
Qy		1444	PWSTGLDGYTGSEOTKDGTSOSSOOKGFDDDFIOALGLKNTVEHGKLGSLGIRIFDPGN--	1501
Dd		1334	---DLDTKLKAHELQSEDVSRDHEKOTYRTLMEEI-----ESLKKELOIFTKANSS	1381
Qy		1502	-----BLAKIKDASNKKGBEKLKSVDLFPKNLYNEYEKSPKIAKGWTNIHP	1548
Dd		1382	SDAFEKLVANNEKEKDRIIDERTKEFEKCLQETLNKSTSEABYSKDIETLKKEWLKEYE	1441
Qy		1549	DQ-----KEYPNPNOQLP--ENYLNIV-----LNQPKVTVLYNSSDFITNLFVEPE	1592
Dd		1442	DETLLRIKEAFENILKGRIRLPSEERIOKIISRKEBELLEEFRKGLKENAGSLTPL-----	1496
Qy		1593	GSDRGSGTKLKVQIQKOVNNNYADWGSAVITFWYDKNIIITNOPNVITANIA-DVFIKDVK	1651
Dd		1497	-DNKGSGEDABEELWNPSKNSERPASAAGFINQKNL---KPQEQLKNVANDVSFNDSQ	1552
Qy		1652	ELENTKLIANIITQWPWNISGSKEBKFKYPTVPFGCN-WENENSMSWSAQOTPTWEKIREG	1710
Dd		1553	SMVNTKE---NNIVD---SSAAGNKA---IPTFSFGKPPFFSNSITS-----	1588
Qy		1711	FALCALKSSFPQKTRTFVLITNAPLMLWKYGP-LGFONGPNFKTQDWRLVFONDNDNIIAA	1769
Dd		1589	-SLAGFONPPTASQSN--INTNAPLRTLINIQPEVAHVAKAI NFSN-----VTDLTNNSTDG	1640
Qy		1770	LRVQE---QDRPBKS---SEDKDKOK	1789
Dd		1641	AKITEIGSTSKRPIESGSTSSPDPTKK	1666

RESULT 20	
AAB18171	
ID AAB18171 standard; protein; 1979 AA.	
XX AC	
XX AC	
DT 07-NOV-2000 (first entry)	
XX XX	
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:28.	
XX KW	
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine; antimalarial; malaria; protozoacide; infection; insecticide.	
XX OS	
XX Plasmodium falciparum.	
PN WO2000025728-A2.	
XX 11-MAY-2000.	
PF 05-NOV-1999; 95WO-US026796.	
PR 05-NOV-1998; 98US-0107131P.	
PA (HOFF/) HOFFMAN S.	
PA (CARU/) CARUCCI D.	
PA (GARD/) GARDNER M.	

Db	389	-----HNFLHMEQDKLNSFV-----KNNOLKV-----	414	Db	1184	GKKKUKKQDVKKQKIEKLNKLTCKNKQIDBELNBEVEKLNNEIETIITYSNDLNKFKDM	1243
QY	745	YDGEIKAPFGLFLPKELRRSSNGSGSONSPWEQOEIISQFDONLSNQDLAOPS	804	QY	1723	KTRTFVLTNAPLPLWKYGLFGFQNGPNFKTDWRLVFNQDDN-QTAALRVOEQDRPEKS	1781
Db	415	YKCEIKN-----LKTELEKKEKELKOIENVSKKEINKLNQNEK-----EKQILAFN	462	Db	1244	KENNLMMKLD-----ENEDNIKMKKSKIDDMKEIKY	1275
QY	805	TKIWEKIIGDENFDQNNRLQYKLLKDLQESWINKTRDNLNLYTMYLWGDKLKVPKNNLEAK	864	QY	1782	SEDKQKQKW-----JKFKVVIPEEM---FNSQNIIRP	1809
Db	463	KNHKEEIHGLKEELKESVKITKTETQELQEM-----VDIK	497	Db	1276	REDEKRNLINEINNLLKK---NEDMCIKYNEMNIKY	1308
QY	865	FRISNLQELLTAFYTSAL-----SNWNYYQDSGAKSTIIEETAEILDPKVKEKVGAD	919	RESULT 21			
Db	498	QKELDQOEKYNQAIESISIEKSKKEYNQY-----KNTYI--EEINNLEKL--EETNKE	550	ID	ABP73774	standard; protein; 1948 AA.	
QY	920	VYQLKFHYAIGFDNAGKFNQEVIRSSRTIYLKTSKSKLEADTTDQLNQAVKNAPLGL	979	XX	ABP73774;		
Db	551	YTNLQNNY-----TNEINWLND--IHMLNGNIKTWNTQI	583	XX	30-JAN-2003	(first entry)	
QY	980	QSFYLDTERFVGFQKLATSIAVQHOKKEXTLPKLANNDGYTLIHDKLKXPVIPQISSPE	1039	DE	Candida albicans essential protein SEQ ID NO 7611.		
Db	584	STLKNVD-----HLLNEQI--DKLANNEKGL--NSKISELNVQIMDLKEE	624	XX	Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;		
QY	1040	KDWFEGL-----NQGSQNVNVSFTGSIIESPYFSTNFOE	1076	KW	signal transduction; DNA replication; cell division; growth;		
Db	625	KDFLNNQIVDLSNOIDLLTRKMEKENKMLEQENKYQEMELLR--GNIKSENILNDEE	683	KW	proliferation; Candida albicans; fungicide; antifungal.		
QY	1077	DADLDQGDQDSRQGNNSLDNAGLLKQ-----KLAILGNQFIQYQONDKEIEPEIIN	1132	XX	Candida albicans.		
Db	684	VCDLKR-----KLSKESKWKMBEHDKKLAELKDDCDVRIRENNEK-----N	727	XX	WO200253728-A2.		
QY	1133	VEKVSLSRVEPKLAKTLEDNGKTIIRVLSDETMSLIVNTTIEKTPMSAVPEVFTKWV	1192	XX	11-JUL-2002.		
Db	728	EDKINMLKEBEDKINTLKEQNEDKINTLKEQNEKINTLKEYEKHKINTMKEEYEHK--	785	XX	26-DEC-2001; 2001WO-US049486.		
QY	1193	EQYDPRPLAAKTKFVLKFDQIPVDGSGNISDKWLASIPLVHQOQLRLSPVKTIRE-	1251	PR	29-DEC-2001; 2000US-0259128P.		
Db	786	-----INTLNEQ	792	PR	20-FEB-2001; 2001US-00792024.		
QY	1252	-----LGLKTEQQOQQOQQOQQOQKAVRKEBELETYNP--KDEF--NILNPLTKAHLRTL	1305	PR	22-AUG-2001; 2001US-0314050P.		
Db	793	NEHKINTLNEQNEHKINTMKEEYEDKQNTLNEQNEDEKMSLKEEYENKINQINSNEIKI	852	XX	(ELIT-) ELITRA PHARM INC.		
QY	1306	SNLVNNDPNY--KIEDLKVI--KNBAGDHQAFS--LRANNIKRLMNTPI-----T	1350	XX	Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;		
Db	853	KDVVN---EYIEBVDKLVTLDEKKQFQKEINYAHIKAHKEQIILLTEMEELKCORDNK	909	XX	WPI: 2002-566694/50.		
QY	1351	FAD-VNPPF-----YYNED-WRSIDKYLNNKGNVSSHQQOQAGNQS	1391	XX	N-PSDB; ABZ32324.		
Db	910	YSDLYEKYIKLKIOWINIBECDDIENEDIIRIEEYINN-----	951	CC	Constructing strains for identifying gene products as effective targets for therapeutic intervention, by inactivating in the strain one allele of a gene and placing other allele of the gene under conditional expression.		
QY	1392	GLIQRLNKNIKPTFTPALIALKDRNNTLSNYSDKIIMIKPKYLVERSIGVPSWSTGLDG	1451	CC	Claim 44; SEQ ID NO 7611; 167pp + Sequence Listing; English.		
Db	952	-----NKGLKE-----VVEKEKRRHSSN--ILKSKEF-----	979	CC	The invention relates to constructing (M1) a strain of diploid fungal cells in which both alleles of a gene are modified, comprising modifying one allele by insertion or replacement by a cassette having an expressible selectable marker and modifying other allele by recombination, of a promoter replacement fragment with a heterologous promoter, so that expression of the second allele is regulated by the promoter. (M1) is useful for constructing a strain of diploid fungal cells in which both alleles of a gene are modified. The diploid fungal cells having both alleles modified are useful for identifying a gene that is essential to the survival or growth of a fungus, a gene that contributes to the virulence and/or pathogenicity of a fungus, a gene that contributes to the resistance of a diploid fungus to an antifungal agent, an antifungal agent that inhibits the growth of a diploid fungus and for identifying a therapeutic agent for treatment of a mammalian disease. (M1) is useful for identifying a compound which modulates the activity of a gene product, preferably enzymatic activity, carbon compound catabolism, biosynthetic, transporter, transcriptional, translational, signal transduction, DNA replication and cell division activity. The method is useful for identifying a compound having the ability to inhibit growth or proliferation of C. albicans cells and for treating infection by C. albicans. The present sequence is that of an essential Candida albicans protein used in the method of the invention.		
QY	1452	YIGSEOTKGTSSSSQKGFQDFIOALGLKNTVEYHGLSLRIPDPGNEIAKIKD----	1508	CC	Note: The sequence data for this patent is not represented in the printed		
Db	980	FKNSIEDK-----SHELKXKHEKDLKS---KDKIEBEK---NKKIKELNNDIKLQDEIL	1028	CC			
QY	1509	-----ASNKGEELKLSYDLFKNYLNEYEKSPKIAKGWNTIHPDQKEYPNPQKLPENY	1564	CC			
Db	1029	VYKQSQAOQVHHKSKSWILLK-----DKSKEKIDKENQINVEKNEEKOLKCKDDE--	1080	CC			
QY	1565	LNVLNLQ---PWKVTLYN-----SSDFITNLFVEPEGSRGSGTKLQVIQKQVNNN	1613	CC			
Db	1081	-IRILNEELVKYKTYILNKLKDPQLLQNDLLSKI-----DINSITINEGCMVDKIEEH	1132	CC			
QY	1614	YADWGSAYLTFWYDKNIITNPQNV-----ITANIADVFIKQVKELEDNTK--LIAP	1662	CC			
Db	1133	ILD-----YDEBKRSRSLFQKNEICSLTTEVMELNKKNELTEENKMLVDQ	1183	CC			
QY	1663	NITQWPNISGSKYKYPVTFPGFNWENENSSMNSQAQTPTEKIREGALQALKSSPDQ	1722	CC			

CC	specification but is based on sequence information supplied to Derwent by			
CC	the European Patent Office			
XX				
SQ	Sequence 1948 AA;			
	Query Match	3.0%; Score 288; DB 5; Length 1948;		
	Best Local Similarity	19.3%; Pred. No. 1.8e-07;		
	Matches 348; Conservative	298; Mismatches 676; Indels 484; Gaps 90;		
Qy	9	LLATAAAIIGSTVFTGVVGLASKYKRGVNPQTQGVISQGLIDSVAFKPSIANFTSD---	65	
Db	339	LLVEAFKIMGFT-----TNKNPF--IYQVLAIIHLGNITFTSNKSEQANTDNDSP	388	
Qy	66	-----YQGVKALLNGKTFDPKSEFTDFVSKDFLTNGRTVLEIPIKKYQVVISSEPS	120	
Db	389	DRICELAAVDKDLFVQNLRLPKVAGREFITK-----SKKNEVKFAIDAFAK-	436	
Qy	121	DKERFRGLFHLEKLEGDGNIAGSATFIYLLPLDMPKAAALQYSYIVDKNFNLIHPL	180	
Db	437	-----YLYEKLFOFIITKINEK-----LDQSHNGDDNTANNNDNFFIGVL	478	
Qy	181	-----SNFSAQSIKPLATRSDFIAKLNQFNQDELWVYLEKPDFDLEALKANIRLOTAD	235	
Db	479	DIAGFEIIFINSFEQCLINYNE--KLQQFFHHSF--ILEQ--SEYLRNINWEIFID	530	
Qy	236	FSFEKGNLVD-----PPVYSFIRNPQKQWASDLNQDQKTIVRLYLRTESPOAKTILK	289	
Db	531	FGQDLQPTIDLTETKQPMGILKLDBECLMPKSSDASFMEKLSKNFTNTH-----	580	
Qy	290	DYKYKDETFLSSIDLKASNGTSLPANENDLKQDLD-----VDLLD-----VSDYF--	334	
Db	581	-KKFSENKFGNGFIHHVAGKVTYVNWLNQNTDPISESLNLLPDSQNELIADMFIND	639	
Qy	335	-----GGQSETITSNSQVVPVPSERSLSKRVKPKDQOK-----PRIEK	374	
Db	640	PHINRPQTNGG-----GNSKLK--TASQKH-KDQLKTLMBQLESTEPHFVRCILPNLEK	690	
Qy	375	FSLYEYDALSFYSQLOELVSPNSIKOLVNATLARNLRFSLGKY--NFLPDDLASHLDYY	432	
Db	691	-RANKFDKNLVGLQR-----CNGVL-EGIRITRAGYPNRMFMDFEP---IQRY	733	
Qy	433	FLVSKAKIKQSIITKFLPIELPIKLSKSSILGDQEPNI---KTLFEKEVTFKLDNFRDV	489	
Db	734	SIICUNEL--SSPQNTNCE-----IILAPVLNPEDFKVLTKIFFKNGILGKLEIRDL	787	
Qy	490	ETEKAFGLLYPGVN-----BELEQAKAQRASFEEKSKG--LKPSFOOK-	533	
Db	788	ALKNIFTDLQKVIKGNLRLVLKQKIKEIQSAQIISRTWVTLDEIKSNSPWWRLPFHVKP	847	
Qy	534	-----EENSKAINN--QEGLE-----EDDNITERL--PENSPLOYQOEN-----AGL	571	
Db	848	LLEDQAKVLDSKGLQENLQTLTVKLKDSFKLTGLETDEKLEKREQMNLDQEMINITSIA	907	
Qy	572	GASPKPYMKDVQORYYLAQSOQL- IKAKDVTYKLAKLSNRHTYNISIRLKEQLFD	630	
Db	908	KEKDKLSQLRUTENK-----SKHRIETLEIKLDFEKQNDLINEHE-----KLTRRSLE	958	
Qy	631	VNPRIPSSRDIEKAFVLDKTEKNKYQIYSSASPVFQNKWSLFGYRYRLLGLDLPQTIH	690	
Db	959	LH-----DKINAKV-----EELNSLNLKLDHSAQSELSNL-----KQOLK	992	
Qy	691	ELVKGQKAGLOFEGVENLPSDFNLEDLK-----NIRIKTPIFPOKD-----	732	
Db	993	EITKLQNAHSEKLLPKQW-HDKSTEDLNEKVSRLLENLKKLQADLNSKDVISSEHGKL	1051	
Qy	733	-NFKLSLLDFNNYDGEIKAPE-----FGLPLFLPKELRRNSNGSGSONSPWEQEI---	785	
Db	1052	QNEIMKLELCSDYERKQKDSQRTTDSLQIQUKKO-----EDYKSYEKIEEA	1100	
Qy	786	-----ISQFDQNLSNQDLAQFSTKIWEKIIIGDENEFQNNRLOYKL	828	
Db	1101	REKVTLKGVKVKSEIQYKSEIKKKAELASSNTKI-----TQLDQSTK-ELSS	1151	

Qy	829	LKDLOESW--INKTRDNLWYTVLGDKLKVKPKNNLEAKFR-----QISNLOELLTAFYTS	881	
Db	1152	LKSNESKYVAETESAKQOL-----SKLIEEVSIEADYNRLQTELKQMKQNTVEYSTR	1204	
Qy	882	-AALSNNWNYQDSGAKSTI-----IPEEIAELDPKVKKEKVGADYVQLKFHYAIG	930	
Db	1205	ITELSTKLQNVTEBAKSKBIKENQPPNPFQMBEFTHMKLVKNEQT-ASLRKEKFENK-K	1262	
Qy	931	FDNAGKFNQEVIRSSRTIYKTS-----GSKLEADTIDQLNOAVKNAPLGLQS FYL	984	
Db	1263	LSBELOQLKBERINMGSLTSMDLTPKRKSLAIGDKSMITNTVDSFNKEIENLKFQLO---	1318	
Qy	985	DTBRFGVQFQKATSLAVQHKQKEKTLFKLNNNDGYTLIHDKLKPKVIPQIPISSPEKDWFE	1044	
Db	1319	--QEQGNFQR-AENYAI---ELQKLNKLLTTRGLNTND-----YE	1354	
Qy	1045	GKLNQNGSQSNVNVSTFGSIIESPYFSTNFQEDADLDQDQD--DSRQ-----GNNSLDN	1097	
Db	1355	KTYNDSQKRITQLETKITGRLLAN---SGGDYEDHLLDDDDWESSRNSGGSGGSSSS	1411	
Qy	1098	QEAGLLKQKLAILLGN-----QFIOYQOQNDKEIEF--EIIINVEKVSSELSFRV----	1144	
Db	1412	ARNSLVKSESITAFPHNMRGVSQDIYIYQDITTKLSSTREELNGSKTEILRLKALLRESE	1471	
Qy	1145	FKLAKTLEDNGKT-IRVLSDETMSLIV--NTTIKTEPMSAVPEVDFTKWVEQVDRPTPL	1201	
Db	1472	DELYQVQKQNYKTSVHDYEQDLAQLKVKCHETLSRNKDINESLEIY-KKRSDEYKKLEL	1530	
Qy	1202	AAKTKFVLK-FKQDIIPVDGSGNISDKWLASIPLVIHQOQLRLSPVVTIIRBELGKTFEQQ	1260	
Db	1531	AEQAIAISKRHEBQATKEMKESRSQ-----LLLVRRELRITQIL--IKDPRIKVENLE	1581	
Qy	1261	QOQOQOQOQOQKAVRKEELETYNPKDFNLIILN---PLTKAHLUTLNLVNDPNYK	1316	
Db	1582	ATIEEKHQLDANK-----BEIKQIQDKLNLHLKFNENKELKEEIKNL-NRDLDPK	1635	
Qy	1317	IE-DLVKIKNEAGDHQALAFS---LRANNI-----KRLMNTPIITFADYNPPFYNN	1361	
Db	1636	TDIETKLK-ENKGLQLDYEDVLLVKNLQNEVEELILQEEKLN-----	1679	
Qy	1362	EDWRSIDKYLNNKNGVSSHQOAGGQSGSLQRLNKNIKPETFTTALIALKDRNNTNL	1421	
Db	1680	---KIDELTNNRNLEN-----EKLINERKIVNCTKQISGLKE-----	1714	
Qy	1422	SNYSDKLIIMIKPKYLVERSIGVPWSTGLDGVISGEQTK-DGTSSSSQKGFDDQFIQ---	1477	
Db	1715	--LVDEISIERDKLLDKKET-----LQNDLQTMKTFDATTTELKQAHGELDFLKXHL	1765	
Qy	1478	-----ALGLKNTVEYHGKLGSLIRIFDP---GNELAKIKDAS---NKKGEKLLKSYDL-	1524	
Db	1766	ENQREDSEAIKTELNQSKMSTSPDIRQOQKLRELLVTKBEENFSLVTKNELNLKVSdle	1825	
Qy	1525	PKNYLNE-----YEKKSPKIAKGTWNIHPDQKEYPNQKLPENYLNL---VLNQPKWYTL	1577	
Db	1826	EKLYSNEQLKYWESKVDTLKALDGLALNEKHEADTKIKLQRSIKQLEIRVENESQSKR	1885	
Qy	1578	YNSDDF	1583	
Db	1886	YNDENF	1891	
RESULT 22				
ADB75595				
ID	ADB75595 standard; protein; 1979 AA.			
XX				
AC	ADB75595;			
XX				
DT	04-DEC-2003 (first entry)			
XX				
DE	Prostate cancer marker protein.			
XX				
KW	Prostate; cancer; cytostatic; gene therapy; marker.			
XX				

OS Homo sapiens.
XX WO2003009814-A2.
XX 06-FEB-2003.
XX 25-JUL-2002; 2002WO-US023913.
XX 25-JUL-2001; 2001US-0307992P.
XX 25-AUG-2001; 2001US-0314356P.
XX 25-SEP-2001; 2001US-0325020P.
XX 12-DEC-2001; 2001US-0341746P.
XX 05-MAR-2002; 2002US-0362158P.
XX (MILL-) MILLENNIUM PHARM INC.
XX Schlegel R, Monahan JB, Endege WO, Gannavarapu M, Gorbacheva B;
PI Hoersh S, Kamatkar S, Wonsey AM, Giatt K, Zhao X, Anderson D;
XX WPI; 2003-248033/24.
XX New nucleic acid molecule, useful for diagnosing or treating prostate
PT cancer.
XX Disclosure; SEQ ID NO 419; 99pp; English.
XX The invention relates to newly discovered cancer markers associated with
CC the cancerous state of prostate cells. Also disclosed is a method of
CC assessing whether a patient is afflicted with prostate cancer. The method
CC of the invention involves assessing whether a patient is afflicted with
CC prostate cancer by comparing the level of expression of a marker in a
CC patient sample and the normal level of expression of the marker in a
CC control non-prostate cancer sample, where a significant increase in the
CC level of expression of the marker in the patient sample and the normal
CC level indicates that the patient is afflicted with prostate cancer.
CC Nucleic acids of the invention are useful for diagnosing or treating
CC prostate cancer, and may be useful in gene therapy. Sequences given in
CC ADB75177-ADB75631 represent marker cDNA and proteins. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1979 AA;
SQ
Query Match 2.9%; Score 286.5; DB 7; Length 1979;
Best Local Similarity 19.0%; Pred. No. 2.2e-07;
Matches 373; Conservative 345; Mismatches 682; Indels 563; Gaps 99;
22 FGTVGLASKVYRGVNPQTGVLISQLGLDSDVAFKPSIANFTSD-----YQSVKKAL 73
164 FGDIISSQOEIN-RLSNEVSRLESEVGHWRHIA-OTSKAQGTDSQSEICKLQNIKEL 221
74 LNKGTDPKPSSEPTDFVSKPDLFTN-NGRTVLEIPKKYQVVISSEFSPEDDKERFLGPHL 132
222 KQN-----RSQIEDDHQHEMSVLONAHQKLTISRREELSDY-----EERI----- 265
133 KEKLEPGNTAQSATKIYLLPLDMPKAAQLGQSVYIVDKNFNNLIHPLNFAQSQIKPLA 192
266 -EELE--NLLQOGSGV--IETDLSKI-----YEMQKTIQVLIQEKVE--STKKMEQLE 312
193 LTRSSDFIAKLNQFN-----QDELWVYLEKFFDLEALKANIRLQ-----TA 234
313 -DKIKDINKKLSAENDRDILRREQSQLNV--EKROIIMECE--NLKLECSKLOPSAVKOS 368
235 DFSFEKGNLV-----DPP-VYSFIRNPQNKWEASDLNQD-----QKTVRLYLRTEFSP 282
369 DTWTEKERILAQASAEVEFRLLQALSDAENEIMRLSSLNQDNSLAEDNLKMKRIEVL 428
283 QAKTILKDYKYDETFLSSIDLK-----ASNGTSLFANENDLKQDLVDLDDVSDY 333
429 KEKSLLS--QEKELOMSLLKLNNEYEVIKSTATRDISLSDSELHDLRLNLE----- 477
334 FGQSQSETITSNSQVQVPASERSLKDVRVKFKDQKPRIEKFSLSYEDALSFFYSQLELV 393

Db 478 -----AKEQLNQSGIS-EKETLIAEIEELDRQNEQATKMILTKQQL 518
QY 394 SKPNSIKDLVNATLARNLRFSLGKYNFLPD---DLASHLDYVFLVSKAKIKOSSITKCLF 450
Db 519 SKQONEGDSIIISKLKQDLNDEKRVHQLEDDKWDITKELD---VQEKLIQSEVA--- 570
QY 451 IELPTKISUKSSILGQDEPNIKTLPEKVTVPKLDNPRDVEIEKAFGLLYPGVNEEUEQAR 510
Db 571 -----LNDLH-----LTQKLEDKVENL-----VDQLN 593
QY 511 KAQRASFEKEKSKGLKEPSQOKEENSKAINNQ--EGLEEDDN-----ITER----- 555
Db 594 KSOENSVSIQENLELKEHIRQNEBELSRIRNELMQSLNQDSNSNPKOTLLKEREAEVRN 653
QY 556 LPEN-SPIQOQENAGLGPDPKPYMIKDVQORVYLA---KSQIQELIKAKDYTKLAK 610
Db 654 LKONLSELEQLNENLKKVAPDVK-----MENEKLVLACEDVRHQLEECLAGNQSLSEK 707
QY 611 LLSNRHTYINISURLKEQLPDVNPRISSRDIEKAPVLDKTEKNKYWQYIYSSASPVFQNK 670
Db 708 -----NTIVETLMEKGEIEAELCWAKKRLLEA-----NKYEKTIEELSNA-RN- 751
QY 671 WSLFGYRYRLGLDPKQTIHE-LVKLGOKAGLOF-----EGYENLPSDF-----NLE 716
Db 752 -----LNTSALQLEHEHLIKLNQKKOMEIAELKKNIEQMDTDHKETKDVLSSSLE 801
QY 717 DLKNIRITPLPSQKDNFKLSLLDNFNYYDGIKAPFGLPLFLPKELARN----- 767
Db 802 EQQL---TQLINKKEIFIEKLERSSKLEELDK-----YSQALRKNEILRQTIEE 850
QY 769 SNSGGSQNSNPWEIISQPKDQ-----NLSNQDLA-----QFST----- 805
Db 851 KDRSLGSKKEENHHLQEELERLERESQRTAPVADPKTLDSTVTELASSEVSLNTIKEHLEE 910
QY 806 --KIWEKIGDNEFDQNNRLQYKLLKDLQSSWINKTRDNLVYTLGDKLVKPKNNLEA 863
Db 911 EIKHKQIIEEDNQ---SKMQ--LLQSLQEO--KKEMDE--FRYQHEQMNATHQTLFLE 960
QY 864 KFRQISNLQELLTAPYT-----SAALSNNMNYQDSGAFTI 900
Db 961 KDEEIKSLQKTIQIKTQLHEERQDIQTDNSDI FQETKVQSLNIENGSEKHDLSKAETER 1020
QY 901 IPEETAELDPKV---EK---VGADVYQLKHYAIGFDDNACKFQEV----- 942
Db 1021 LVKGIKERELEIKLNEKNISLTKIDQLS-----KDEVKLTQIIQOKOLEIQALHA 1073
QY 943 -IRSSSRT---LYLKTSGSKLEADTIQO-----LNQAVKNAPLGLOSYLDTERFVGF 992
Db 1074 RISSSTHTQDVVYL---QQQLQAVAMEREKVFVAVLNKTR-----ENSHLKTVEHKM 1123
QY 993 QKLATSLA--VQHKQKERTLPKLLNDGVTYLHDKLKKPVIPOISSSPKDFEGKLNQ 1050
Db 1124 DIVAAKEAALIKLDENKKLSRPFSSGQDMFRETION--LSRIIR--EKDIEDALSQK 1179
QY 1051 GQS-----QNVNVSTFGSIIIESPYSTNFQEDADLD-----QDQDSDS 1088
Db 1180 CQTLAVLQTSSTGNEAGGVNSHQFELLQERDKLQVKKMEWKKQVMTTVOQNMOHES 1239
QY 1089 RGNNSLNDQEAAGL---KOKLAILLGNOPFYVQQNDKEI----- 1126
Db 1240 AOLQELHQLQAVLVSDNNNSKLQVDY-TGLIQSYEQNETKLKNFGQELAQVHISIGL 1298
QY 1127 -----EPEIIN-----VEKVSLSFRVEFKLAKTLSD 1153
Db 1299 CNTKDLLLKLKLDIISPOLSSASLLTPQSAECLRASKSEVLSSELSLQOELBELRKSLOE 1358
QY 1154 NGKTRIVLSDE---TMSLIVNTTIEKTPENSAVEPDTKWVEQVDPPTPLAAKTFVL 1209
Db 1359 KDATIRTLQENNRHRLSDIAATSELERKEHEQTDSEIKOLK--EKQDVLQKLKELDLI 1416
QY 1210 KFKDQIPVDGSGNISDKW-----LASIPLVHQMLRL-----SPVVKTIRELG 1253

Db 1417 KAKSQDSSNNFTNKNVENELLRQAVTNLKERILILEMDIGKLGKGENKIVETRY--G 1474
Qy 1254 LKTEQOQOQOQOQOQOQKAVKKEBELETYNPKDFNINLPUTKAHRLTSLNVLNNDP 1313
Db 1475 KETEYQALQETNMKFSM-----MLREKEFECHSMKALAPEQLLKEQKGTGELNQLL 1529
Qy 1314 N--KYIEDLVKIKNEAGDHQALAFSIRANNIRKLANTPTTFADYNFFFYNNEDRSIDKYL 1371
Db 1530 NAVSMQKTVVFOQERD-QVMLALQKQ-----KOMENTAL-----ONEVQRLDK-- 1573
Qy 1372 NNGKNVSHQOQAGGNGSGLIORLNKNI--KPEFTFPALIALKDRNTNLSNYSKII 1429
Db 1574 ---EPRNQE-----LERLRNHLESSEDSYTRALAAEDRE---AKLRKKVT 1614
Qy 1430 MIKPKYLVRSIGVPMSTGLDYGIGSEQTKDGTSSSQKGFDDQDFTQ-----A 1478
Db 1615 VLEER-LVSSSSNAME-NASHQASQVQESLOBLQNLVSVKQR--DETALQLSVSQBQVQYA 1670
Qy 1479 LGLKNTYHVGKLGSLIRIFDP-----GNEIAKID--ASNKKGEELKLSYDLFKNYLN 1530
Db 1671 LSLAN-----LQWLEHFOQEEKAMYSAELEKQKOLIAEWKKNAENLEGKVISLQECID 1724
Qy 1531 EYEKSPKIAK--GWTNIHPQKQEPNPNQKLPENYLNVLNQPWKVTLNNSD----- 1582
Db 1725 EANAALDSASRLTEQLDVKEQIEELKEQNELRQEMLDV--QKKLMSLANSSEKVDKV 1782
Qy 1583 FITNLFV-----EPEGSDR-----GSGTKLKQVIOQVNNVYADWGA--VLTFWY----- 1626
Db 1783 LMRNLFIGHFTPKNQREVLRLMGSILGVRRREMEQLFHD--DQGSVTRWMTGLMGGS 1840
Qy 1627 ----DKNIITQPNVITANIADVFKDVKLEEDNTKLIAPNIT 1665
Db 1841 KSPVNTPLRPNQSVVNSFSSELFVKFL-ETESHPSPPPKILS 1882

RESULT 23

ABU35589
ID ABU35589 standard; protein; 1805 AA.
AC ABU35589;
XX
XX 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #21116.
DE Antisense; prokaryotic essential gene; cell proliferation; drug design.
KW Mycoplasma genitalium.
OS
XX WO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA Wang L, Zamudio C, Malone C, Habelbeck R, Ohlsen K, Zyskind JW;
PI Wall D, Ttawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
PI
XX WPI; 2003-029926/02.
DR N-PSDB; ACA39459.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX

Claim 25; SEQ ID NO 63513; 1766pp; English.

PS The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

Sequence 1805 AA;

Query Match 2.9%; Score 281; DB 6; Length 1805;
Best Local Similarity 19.2%; Pred. No. 4.1e-07;
Matches 361; Conservative 282; Mismatches 627; Indels 614; Gaps 93;
Qy 77 KTFDPKSS-----EFTDFVSKEDFLTNNGRTVLEIPKKYQVVISFSPEDDKERPLG 129
Db 2 KPFDKPSLOPIYDIGDFDDGYLQSEYEKRSKT--DVDKTIENQLLKEIKSLEDELKNG 59
Qy 130 FHLEKLEDGNIQAQATKFIYLLPLDMPKALQGVSYVDKNFNNL-----IHLPSNF 183
Db 60 --LKNQAEIN-----PE-----LDKKINHLEVDLNRNLVNEYKNF 91
Qy 184 SAQSIKPLALTRSSDFIAKLNQFNNOQDELWYLEKFFDLEALKANIRLQTADFSEK-GN 242
Db 92 QFQ-----KNHWDVKVSELDN-----LTFYKVELTR--LQENADFLNSKYAN 133
Qy 243 LVD-----PFVYSFIRNPQKQEWASDLN-----DQKTVRLYLRTFSPQAKILKDYK 292
Db 134 LANFQANYHNKLNDFHRLIENQNTINLNQKINGNQNLIDNNVALLQNPITVEKKNY- 192
Qy 293 YKDETFLLSSIDLKASNGTSLPANENDLKQDLVD--LLDVSDYFGGQSETTTSNSQKPV 350
Db 193 -----LLNVID-----OLY-----NEL-DQLEENOKRLUSI--EYENTYRELVSADNELQNV 235
Qy 351 PASERSLNDRVKFKDQKQKPRIEKFSLYVEYDALSPYSQLOELVSKPNSIK-----DLV 403
Db 236 --YENIDQNOIQFKHOYQTYRELSQLERKILT-----KQELVDKESALRVKIDDAFVI 289
Qy 404 NATLARNLRFSLGKYNFLFDDLASHLDYVFLVSKAKIKQSSITKFLFELPIKLSKSSI 463
Db 290 NARLAE-----LDDVAKQLSF-----QDGIITKQNAQHVDEKIL-----VA 323
Qy 464 LGDQEPNITLFEKVTPLDNFRDVEIEKAFGLLYPGVNEELEQAKQAASPEKSKS 523
Db 324 LNKEKORLNT--QKEAFFNLRSQALIDINK-----LQOENELFAKHLHQNEFEKQKSD 376
Qy 524 KGLK-----EFSQKBEKSKAINNQE--GLEEDDNITERLPENSPIOYQENA 569

```
Db 377 SLUKLEYKALQHINEFKNESATSEELNQBRELFEKREIDITLLTQASLEYBQR- 435
Qy 570 GLGASDPKPYMIKDVQNRQYLLAKSQTQELIKA-KDYTKLAKLLSNRHTYNISLRKEQL 628
Db 436 -----ESSQLLDKQNE-----VKQHFQNLLEYAKKELDKERNLLDQOKVD-----SEAI 480
Qy 629 FDNVPRISSR-DIEKAKFVLDKTEKNKYWQIYSSASPV-----FONKWSLFGYRYLL 681
Db 481 FOLKEKVAORKELEELYLVKKQKQDQKENELLFFEKQLKQHQADPENE-----529
Qy 682 GLPKQTIHELVLKQKAGIQFEGYENLPDSE-NLEDLKNIRIKTLPFSQKO-NFKLSLL 739
Db 530 -LEAKQ-QBL-----FEAKHALERSFKLED-----KEKDINTKAQOI 565
Qy 740 DFNNYDGEIKAPEFGLPLFLPKELRRNSSGGSQNSNSPWQEIISQFKQNLNSNQDQ 799
Db 566 -----ANEF-----SOLKTDKSKA-----DFELMLQNEYENLQOEKQ 598
Qy 800 LAQFSTKIWKIIGDNEFPQN-----NRLQYKLLKDLQBSWINKTRDNLWYTLGDKLK 854
Db 599 -----KLFQERTYFERNAAVLNRLQKREELQO-----KETLDQLTKSFEQERL- 644
Qy 855 VKPKNLEAK--FRQISNLOELLTAFVYTSAAALNNNNYQDSCAKSTIIFEEITAEIDPKV 912
Db 645 ---INOREHKELVASVEKQKEIL-----GKKLQDPFSQTSLNASKNJAEREMAI 689
Qy 913 KEKVGADVYQLKFHYAIGFDNAGKFNQEVIRSSRTIYLKTSKGKLEADTTIDQLNQA- 971
Db 690 -----KPKKEIEATEKOLLNDVNAEVIQAD-LAQLNQSL 724
Qy 972 -----VKNAPLGQSPYLDTERFGVQKL-ATSLAVQHKQKE-KTLPKKKLNNDGVTLLH 1023
Db 725 NOERSELQNAKORIADPHNDS-----LKKLNEYELSLQKRLQELQLEANKQHSY---- 775
Qy 1024 DKLKPVIPQISSPEKDFEG-----KLNQNGQS-----QNVNVTFGSIIESPYFTN 1073
Db 776 -----QNAVYFEGELDKLNRKQKQFAPLNRKQKQTEVDIAIKORUSDHQAUN 821
Qy 1074 FQEDADLDQDQSDRGNNSLNQBAG--LLKQKLAILLGNQFI-----1116
Db 822 MQQ-AELDR---KTHELNNAFLNHDADQKSLQDLQATVKETQKLDLERSALLEKQREF 876
Qy 1117 -----QYQONDKEIEFEIL--NVEK---VSELSFRVE 1144
Db 877 AENVAGFKRHSNKTSLQKIYELTKKQSEQTKETELKIAFSDLQKQYQVPELQKQOE 936
Qy 1145 FK-----LAKTLEDNGKTIKRVLSDETSLLVN--TTIEKTPMSAVPEVDFTKWE 1193
Db 937 FRQIEAKQRELDKLAENQO-VKLELDNRFOALQONQKQDTVQAOLEL-----E 985
Qy 1194 QYD---PRTPLAAKTKFVLKPKDQIPVDGSGNISDKWLA-SIPLVTHQQMLRLSPVVKTI 1249
Db 986 QHQLNLEQTAFNAQANESLLKQREQ-----LTKKIQAFHYELKKNQFLALKG---- 1032
Qy 1250 RELGLKTEQOQOQOQO-----OQOQOQPKKAVRKE 1279
Db 1033 KRLFAKEQDQQRKQDQINWRPFQKEKYTDDEAKKRELEEKIRRSLSQSNVELEKR 1092
Qy 1280 BELETYNPKDEFNLNPL---TKAHLRTLSNLV-----NNDPNYK-----1316
Db 1093 EKLAT-----DFTNLKNVQNTQINQDLNSQTRQFLERKNRQFNSNANAKAFILKR 1147
Qy 1317 -----IEDLVKIRNEAGDHQALFASLRANNIKR-LMNTPIITPADYNPPFYNEOWRSIDKYL 1371
Db 1148 LRSFASNLKQKEALAIQKLEFKRDEQKQKELQATLQLEQF-----KFE 1193
Qy 1372 NNGKGVSSHQQAAGNGQSGLLQRLNKNIKPTFTFPALIALKDRNNTNLSVSDKIIMI 1431
Db 1194 KQNFIDEKQRLQVAIKTQ-----CEKLSDEKK--ALNQKLVELKNLSQTYLANKN-----1241
Qy 1432 KPKYLVERSIGVPWSTGLDGYIGSEOTKDGTSSTSSSQKGFQDQFQIALGLKNTYHGKLG 1491
```

```
Db 1242 KAEY-SQOOLQOQYTNLLDLKENLERTKD-----QLDKKHSIFARLTKFANDLREKQ 1295
Qy 1492 L--SIRIFDPGNEELAKIKDAS-----NKKGEKLLKSYDL 1524
Db 1296 LLKAQRIVDKRNLLKENERNLHFLSNETERKRAVLEDOIYFEKQKQATDAILASHKE 1355
Qy 1525 FKQYLNVEYKSPKIAKGTWNIHPQOKEYPNPQKLPENYLNVLNQPWKVTLYNSSDFI 1584
Db 1356 VKKGEGLQKLVELETRKTKLNNDFAKFSRQREBFENQRLKLELQKTLQOTNSNPFK 1415
Qy 1585 TNLFVEPEGSDRGSTKLQKQVIQKQVNNNYADWGSAYLTFWYDKNIITNQPNVITANIAD 1644
Db 1416 TKAIQEIENSYKRGWEELN--FQKKE-----FDKN-----KSLYE 1449
Qy 1645 VFIK---DVKELEDNTKLIAPNITOWPNI---SGSKEFYKPTVFFGWNENESSMNSQ 1698
Db 1450 YFRKMRDEIERKESQVKLVKE-TQKANLLEAQANKLIEKNTIDFKE-----1499
Qy 1699 AQPTWEKIREGFAQALKSSFDQ 1722
Db 1500 -----LKAFKDKVDQ 1509

RESULT 24
ABO23572
ID ABO23572 standard; protein; 1805 AA.
XX
AC ABO23572;
XX
DT 04-SEP-2003 (first entry)
XX
DE Mycoplasma genitalium outlier protein #3.
XX
KW Candidate protein identification; pathogen; anti-infective;
KW outlier protein; virulence protein; antigen; drug target protein;
KW pathogenic organism; antimicrobial.
XX
OS Mycoplasma genitalium.
XX
PN US2003039963-A1.
XX
PD 27-FEB-2003.
XX
PF 30-MAR-2001; 2001US-00820843.
XX
PR 30-MAR-2001; 2001US-00820843.
XX
PA (BRAH/) BRAHMACHARI S K.
PA (RAMA/) RAMACHANDRAN S.
PA (NAND/) NANDI T.
PA (BHIM/) BHIMARAO C.
XX
PI Brahmachari SK, Ramachandran S, Nandi T, Bhimarao C;
XX WPI; 2003-492159/46.
XX
PT Identifying candidate proteins useful as anti-infectives involves
PT matching outlier protein sequences with protein sequences in databases.
XX
PS Example 7; Page 70-74; 117pp; English.
XX
CC The present invention relates to a method for identifying candidate
CC proteins in pathogens useful as anti-infectives. The invention discloses
CC a computational method which involves the calculation of several sequence
CC attributes and their subsequent analysis results in the identification
CC of outlier proteins in different pathogens. The method is useful for the
CC identification of outlier proteins (e.g. virulence proteins, antigens or
CC proteins used as drug targets) in pathogenic organisms. The method of the
CC invention provides reproducible results as it does not depend on the
CC variable biochemical characterisation of proteins. ABO23500-ABO23617
CC represent outlier proteins identified from different pathogenic organisms
XX
SQ Sequence 1805 AA;
```

Query Match		2.9%, Score 281; DB 7; Length 1805;
Best Local Similarity		19.2%; Pred. No. 4.1e-07;
Matches		361; Conservative 282; Mismatches 627; Indels 614; Gaps 93;
Qy	77	KTPDKSS-----ETDFVSKFELTNGRTVLEIPKYYQVVISFSPEDDKERPLG 129
Db	2	KPFDPKPSLOPIYDIGDDGVQSEYKNSKT--DVDKIENQLLKIKSLEBELKNLKG 59
Qy	130	PHLKEKLEDGNGIAQSATKFIYLLPLDMPKALGOYSYIVDKNFNL-----IIHPLSNF 183
Db	60	--LNQAEADN-----PE-----LDKKINHLEVDLNRNLVNEYKNF 91
Qy	184	SAQIKPLALTRSDPTAKLQNFQNDDELWYLEKFPDLEALKANILQTDADSFKEK-GN 242
Db	92	QFQ-----KHMVDVKSELND-----LTFYKNELTR--LOENADFLNSKYAN 133
Qy	243	LVD-----PFVYSFIRNPONKEWASDLNQ-----DOKTVRLYLRTBFSQAOKTILKDYK 292
Db	134	LANFOANYHKNLNDPHRLIENQNOTINRLNOKINGNQLNDNNVALLQNFITVEKQNY- 192
Qy	293	YKDETFLSSIDLKASNGTSLFANENDLKQDLVD--LLDVSDFGSGSETITSNQKPV 350
Db	193	-----LLNVID-----QLY--NEL-DQLENQKRLLSI-EYENTYRELVSADNELQNV 235
Qy	351	PASBSLKDRVKFKDQOKPRIKPSLYEYDALSFYQQLVSKPNSIK-----DLV 403
Db	236	--YENIDQNIQFQHQYQTYRDELSQLERKTLT-----KQELVDKESALRVKIDDAFYI 289
Qy	404	NATLARNLRPSLGKYNFLFDLASHLDYVFLVSKAKIKQSSITKKLPFLPIKISLKSSI 463
Db	290	NARLAE-----LDDVAKQLSP-----QDQITQNAQHVEDKL-----VA 323
Qy	464	LGDOEPNITLFEKEVTPLKDNFRDVEIEKAFGLLYPGVNEBELQAKAQRASFEKSKS 523
Db	324	LNKEKORLNT--QKEAFNLQASALIDNK-----LOQENELFAKLEHQNEPEQKQSD 376
Qy	524	KGLK-----BFSQOKEENSKAINNOB-GLEEDDNITERLPPENSPIYOQENA 569
Db	377	SLLKLETKYKALQHKINEFKNESATKSEBELLNORELFEKREIDTLLTOASLYEYHQ- 435
Qy	570	GLGASDPKPYMKDVQORYYLAQSIOELKA-KDYTKLAKLSNRHYNISLRLEQOL 628
Db	436	-----ESSQLLKQKNE-----VKQHFONLEYAKKELDKERNLDDQKKVD-----SEAI 480
Qy	629	FDVNPRISSR--DIEKAKFVLDTKEKNKYQIYSSASPV-----FQNKWSLFGYYRYLL 681
Db	481	FOLKEKVAQERKELEBELYVKKQKQDKENELLPEKQKQHQADFENE----- 529
Qy	682	GLDPQOTIHELKQKAGLOPEGYENLPSPF-NLEDLKNIRIKTPLPSOKD-NFKLSLL 739
Db	530	-LEAKQ--QEL-----FEAKHALERSFIKLED-----KEXDLNTKAQOI 565
Qy	740	DFNNYVDGEIKAPERGLPLFLPKELRRNSSNGSGSONSPWEQIIISQFKDQNLSDQ 799
Db	566	-----ANEF-----SQLTKDKSKA-----DFELMLQNEYENLQEQK 598
Qy	800	LAQFSTKIWEKIGDENEFQDN-----NRLQYKLLKDLQESWINKTRDNLNWTYLGDKLK 854
Db	599	-----KLFOERTYFERNAAVLNRLQKREELLQK-----KETLDLTKSFEQERL- 644
Qy	855	VKPKONLEAK--FROISNLQELLTAFYTSALSNWYNYQDSGAKSTIIFIEAELDPKV 912
Db	645	-----INQREHKELVASVEKQKIL-----GKQLDPSQTSLSNKNLAEREMAI 689
Qy	913	KEKVGADVQLKFAHVAIGDDNAGKFNQBSVIRSSRTTYLTKSGSKLEADTIDOLNOA- 971
Db	690	-----KPEKEITEATEKQLLNDVNAEVIQAD-LAQNLQSL 724
Qy	972	-----VKNAPLGLOSFYLDTERFGVQKL-ATSLAVQHKOKE-KTLPKKNLNDGYTIIH 1023
Db	725	NQERSELQNAQRIADFNHDS-----LKKLNEYELSLQKRLQELQTLLEANQKQHSY---- 775
Qy	1024	DKLKKPVIQIISSPKDWFEQ---KLNQNGOS-----QNVNVSTFGSIIESPYFSTN 1073
Db	776	-----NQAYFEGELDKLNEKQAFNLNRKKQTMVEDVDAIKORLSDKHQALN 821
Qy	1074	FOBDADLDQGDSDSQNGNSLNDNBAG--LLKQKLAILLGNQFI----- 1116
Db	822	MQQ-AELDR-----KTHELNNAFLNHADQKSLQDQLATVKETQKLIIDLSERSALLERQEF 876
Qy	1117	-----OYYQONDKEIFEI--NVEK--VSELSFRVE 1144
Db	877	AENVAGFKRHWSNKTLOLOKIVELTKKQSEQTKETELKIAFSDLOKQOVFELQKQ 936
Qy	1145	FK-----LAKTLSDNGKTIIVLSDETMSLVN--TTIEKTPEMSAVPEVFDTKWVE 1193
Db	937	FRQIEAKQRELDKLAENQK-VKLELDNRFQALQNKQKQDTVQAQLELER-----E 985
Qy	1194	QYD--PRTPLAAKTFVLKFKDQIPVDGSGNISDKWLA-SIPLVIIHQMLRLSPVVKTI 1249
Db	986	QHQLNLEQTAFAFNAESLLKQREQ-----LTKKIQAFHYELKKNQFALQK- 1032
Qy	1250	RELGLKTEQOQOQOQO-----QOQOQOQKQAVRKE 1279
Db	1033	KRLFAKEQOQOQKQDEINWRPKQFEKEYTDFDAKRELEBELEKIRRSLSQSVLEKER 1092
Qy	1280	EELETYNPKDEFNINPL--TKAHLRLTSLNV-----NNDPNYK----- 1316
Db	1093	EKLAT-----DFTLNKVKQHTQJNRDQLNSQIRQFLERKNFQFNSNEANAKKAPLKR 1147
Qy	1317	-----IEDLVKINEAGDHLAFSLRANNIKR-LMWTPIITFADYNPPFYNNEDWRSIDKYL 1371
Db	1148	LRSFASNKLQKEALATQKLEFDRDEQKKELQOATLQLEQF-----KFE 1193
Qy	1372	NKGNVSSHQOQAGGNOGSGLQRLNKNIKPETFTPALIALKDRNNTNLSNYSDKIIMI 1431
Db	1194	KONFDIEKQOLVAIKTQ-----CEKLSDEK--ALNOKLVELKNLSQTYLANK- 1241
Qy	1432	KPKYLVERSIGVPWSTGLDGYIGSEQTKDGTSSSQKQDFDQFIQALGLKNTYHGLKG 1491
Db	1242	KAQY-SQOQLOQKYTNLLDLKENLERTKQ-----QLDKHRSIFARLTKFANDLRPEKKQ 1295
Qy	1492	L--SIRIFDPGNELAKTKDAS-----NKGEEKLLKSYDL 1524
Db	1296	LLKAQRIVDDKNRLKENERNLHFLSNETERKRAVLBDQIYSEFKQKQATDAIASHKE 1355
Qy	1525	FKYLVNEYKSKPIAKGWTNIHPDOKEYPNPKLPENVLNVLNPKVTLNYSDDFI 1584
Db	1356	VKKKEGLQKLLVELETRTKLNNDFAKRSQREEFENQRLKLELOKTLOTQTSNNFK 1415
Qy	1585	TNLFVEPEGSDRSGTKLKQVIQKQVNNNYADMGSAYLTFMYDKNIITNQPNVITANIAD 1644
Db	1416	TKAQBIENSYKRGMEELN--FQKKE-----FDKN-----KSRLYE 1449
Qy	1645	VFIK--DVKELENTKLIAPNITQWPNL-----SGSKEKFKYKPTVFGNWNENSSMNSQ 1698
Db	1450	YFRKWRDEIERKESQVKLVLE-TORKANILLEAQANKLNTKNTIDPFKEB----- 1499
Qy	1699	AQPTWEKIRGEFALQALKSSFDQ 1722
Db	1500	-----LKAFKQKVDQ 1509
RESULT 25		
AAB18161		
ID	AAB18161	standard; protein; 2441 AA.
XX	AAB18161;	
AC	XX	
DT	XX	07-NOV-2000 (first entry)
XX	XX	
DE	XX	Plasmodium falciparum chromosome 2 related protein SEQ ID NO:18.
KW	XX	Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine; antimalarial; malaria; protozoacide; infection; insecticide.

XX OS Plasmodium falciparum.
 XX PN WO200025728-A2.
 XX PD 11-MAY-2000.
 XX PF 05-NOV-1999; 99WO-US026796.
 XX PR 05-NOV-1998; 98US-0107131P.
 XX PA (HOPE/) HOFFMAN S.
 XX PA (CARU/) CARUCCI D.
 XX PA (GARD/) GARDNER M.
 XX PA (VENT/) VENTER J C.
 XX PI Hoffman S, Carucci D, Gardner M, Venter JC;
 XX DR WPI; 2000-365347/31.
 XX PT Proteins encoded by chromosome 2 of the human malarial parasite,
 XX PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 XX PT diagnosis of P.falciparum infection.
 XX PS Disclosure; Page 50-57; 577pp; English.
 XX CC The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (i) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II). (I) and
 CC (II) are useful for the development of vaccines against P. falciparum
 CC infection. (i) and polyclonal antisera or a monoclonal antibody raised to
 CC immunogens comprising the sequences of (I), are useful in the detection
 CC of infection with P. falciparum. Furthermore, (I) (especially when they
 CC are refined or secreted or membrane proteins) can aid the identification
 CC of drugs to treat or prevent P. falciparum infection, or they can be used
 CC to identify drug resistance in P. falciparum. Sequencing of the
 CC Plasmodium chromosome 2 and the subsequent identification of proteins
 CC encoded by it will help to expand our understanding of parasite biology,
 CC a process hampered by the complexity of the parasitic lifecycle, and
 CC provide new targets for vaccine and drug development. Parasite resistance
 CC to drugs and mosquito resistance to insecticides have led to a resurgence
 CC of malaria in many parts of the world, and there is a pressing need for
 CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
 CC represent nucleotide and protein sequences given in the present
 CC invention, but which are not specifically mentioned within the
 XX CC
 XX SQ Sequence 2441 AA;
 Query Match 2.9%; Score 281; DB 3; Length 2441;
 Best Local Similarity 20.6%; Pred. No. 6.5e-07;
 Matches 349; Conservative 234; Mismatches 614; Indels 498; Gaps 85;
 74 LNCK-----TFDP-----KSEFTDFVSKFDFLTNGRTVLEIPKKYQVWVISEFS 118
 481 VNGKLLSTFPFKRRNKLKERMQELHKPKNY-----KKYQKLE--- 523
 119 PEDDKERFLGFLHKEKLE--DGNIAQSA TKFYLLPLDMPKAALGOYSVIVDNKFNLI 176
 524 -----REKREMPDGE-----PLNTPETIHVIRPDLMDKGENKSA 557
 177 IHPLSNFSAQSIK-----PLALTR-SSDFTIAKLQNFNQ--DELWVYL 216
 558 GHFPKPYQTKGLKEYESHVAKDYQLEHEPPTKLPEYKSHVSREYQLDNEVRDELPEYE 617
 217 EKFFDLE-ALKANIRLQTADFSEKGNLVDPFVYSFIRPNQKWEASDLNQDKTVRLY 275
 618 KGHVSREYQLDNEVRDELPE--YEKGHV-----SREYQLD-NEGFPSTLKEY 660
 276 LRTEFSPQAKTILKDYKYKDETFLLSSIDLKASNGTSLFANENDL--KDQLDVLDDVSDY 333
 661 DQELAKGKDTNKPHEVSVEYDQSEL-----AKGKDTNKPHEVSVEYDQTEL 709

334 FGGQSETITSQVKPVPASERSLKDRVKFKDQKQKPRIEKFSLYEYDALSFYSQIQELV 393
 710 AKGKEVTNKPHELE-----EYNETDLAKGKEVTNKPHELE-----ESVDEYDQ-SELAKGKDI 760
 394 SKPN-SIKDLYNATLARNLRF-----SLGKNYFLPDDLA-----SHLDYVFLVS 436
 761 NKPHEVSVEYDQTELAKGKEVTNKPHELEYN--ETDLAKGKEVTNKAHLENLENETD 818
 437 KAKIQSSITKLFTEL-----PIKISLSSILGDOEPNIKTIFE-----KEVTFK----- 482
 819 LAKGKE--VTNKAHLENLEYNETDLAKGKEVTNKAHLENLENETDLAKGKEVTNKAH 876
 483 LDFNRDVEIEKAFGLLYPGVNEELPQARQAQASPEK--EKSXKGLKEFSQO-----KBE 535
 877 LSEYNETDLAKGKEVTNK-ARENLESEYNETDLAKGKEVTNKAHLENLENETDLAKGKEV 935
 536 NSKAINNOGLEEDD-----NITERLPENSPLOYQENAGLIGASDPKPYMIKDVQNO-RY 589
 936 TNKAHLENLEYNETDLAKGKEVTNKAHLENLE- EYNETDLAKG-----KEVTNKARE 985
 590 YLAKGQIOBLIKADYTKLAKLLSNRHTYN-----ISRLKEQLPFDVNPRISSRD 640
 986 NLESEYNETDLAKGKEVTNKAR--ENLESEYNETDLAKGKEVTNKAHLENLEYN-----ETD 1038
 641 IEKAFVLDKTEKN--KYWQIYSSASPVFQNKWSLFGYYRYLLGLDPKQTIHELKVLGOK 698
 1039 LAKGKEVTNKAHLENLESEYEDKYMKNELQNGKS-----DGLKENAELKNK 1084
 699 AGLOEGEYENLPSPDNLELKNIRIKTPLFSOKDNFKLSLLDPNN-----YDGEIKAP 752
 1085 E-LRNKSGDGLKENAELKN-KELRNKGS--DGLKENAELKNKELQNGKSGELKENAELKNK 1141
 753 EF-----GL-----PLFLPKELRNSSNGSGSQNSPWEQEIISQFKQNLNQDLOAQF 803
 1142 ELQNGKSGELKENAELKNKEL-RNKSGSEGLKEN-----AELKNKELQNGKS----- 1186
 804 STKIWEKIIGDNEFPQNNRLQYKLLKOLQBSWINKTRDNLVYLGDKLKVKKPNLSEA 863
 1187 -----EGL--KENAELKNKELQNGKSGELKENAELKNKE-----LQNGKSGELKENAEL 1233
 864 KFRQISN--LOELLTAFYTSALSNNVNYDQSGAK-----STIIPEEIAELDPKVKKEVGS 917
 1234 KNKELRNKSGELKENVYTNNDLKNNDLQNKLSNKMKNKELLNKDISKNMKNKELLN 1293
 918 ADVYQLKFIYAIGFDDNAGKFNQEVIRSSRTIYLTSGSKLEADTIDQLNAQVKNAPL 977
 1294 KDL-----SNEDMKNELLNKDIANKDLKSIG-----NMEQQNTGLKNTP- 1333
 978 GLQSFYLDTERFGVFKLATSIAVQHKKEKTLPLKLLANDGYTLIHDKLKK-PVTPQISS 1036
 1334 -----SKGQNTGLKNTPNERQNTG-----LKNTPSEGOQNT 1366
 1037 SPEKDWFEGLKN-----QNGOSQNVNVTSPGSIIESPVSTNFQDADLDQ---DGQDD 1087
 1367 GUKNTPSEGOQNTGLKNTPNERQNTGLKNTPS-----EGQNTGLKNTPIEGQON 1417
 1088 SPQGNNSLDNQAGLLKQKLAILLGNQFIQYQQYQNDKEIEFEIINVEKVSSELSFRVEFKL 1147
 1418 TGLKNTPSEGOQNTGLKN-----TPNERQNTGLKNAANKGQOQNTGLKNTP 1463
 1148 AKTLSDNGKTRVLSDETWSMLIVNTTIKTPMSAVPEVFTKWVEQYDPRTPLAATKTF 1207
 1464 SKGOQNTG-----LKNTPNERQNTGLKNTP-----NERQNTGL----- 1498
 1208 VLKFKDQIPVDGSGNISDKWLASIPLVIHQMLRLSPVVKTIRELGLKTEQOQOQOQ--- 1264
 1499 -----KNTPSEGOQNTGLKN-----LKNTPNERQNTGLKNTASKGQOQNTGLKNAPNERQNTGL 1550
 1265 -----QOQOQOQOKKAVRKEELET--YNPKDE-----FNILNPLTKAHLRJTLSNLVNDPN 1314
 1551 KNTPSEGOQNTGLKNTPSEGOQNTGLKNTPNERQNTGLKNTASKGQOQNT---GLKNAPN 1607

Qy 1315 YKIEDLVKIVEAGDQLAFSLRANNIKRLMNTPTITADYNPPFYNNEDHRSIDKYLNNK 1374
 Db 1608 EROQNTG-LKNTPEGQNTGLKNSASKGQNTGL-----KNTPEGQNNDL 1654
 Qy 1375 GNVSSHQQAAG-----GNQSGLIQRLNKNIKPETFPTALIALKDRNNLNSYSDK 1427
 Db 1655 KNAPNERQNTGLKNTPEGQNTGL-----KN-----TPS-----SGQNTGLKNTPS- 1698
 Qy 1428 IIMIKPKYLVERSIGVPMSTGLDGYIGSEQTKDG---TSSSSQKGFDPDFIQAIGLKNT 1484
 Db 1699 -----GGQNTGLKNTPEQNTGLKNTPEGQNTGL-----TGLKNT 1735
 Qy 1485 --EYHGKLGISIRIFDPGNEALAKIKDASNKGEEKLLKSYDLFNKYLNEVE-----KKSP 1537
 Db 1736 PNERQNTGL-----KNAANKGQNTGL-----KNTPEGQNTGLKNTP 1775
 Qy 1538 KIAKGWNTIHPDQKEYPNPNQKLPENLVNLVNPWKVLYNSSDFITNLFVEPEGSDRG 1597
 Db 1776 SEGQNTGL-----KNTPEGQNTGL-----NTGLKNT---PSEGQNTGLKNTPNRQ 1818
 Qy 1598 SGTGLKQVIQOVNN 1612
 Db 1819 QNTGLKNAANKGQNT 1833

RESULT 26

ABU36321
 ID ABU36321 standard; protein; 1818 AA.

XX AC ABU36321;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #21848.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Mycoplasma pneumoniae.

XX FN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-0009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR N-PSDB; ACA40191.

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 64245; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of

XX CC the 6213 antisense sequences given in the specification where expression

XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:

XX CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 1818 AA;

Query Match 2.9%; Score 280.5; DB 6; Length 1818;
 Best Local Similarity 19.6%; Pred. No. 4.4e-07;
 Matches 310; Conservative 261; Mismatches 628; Indels 383; Gaps 66;
 Qy 56 KPSIANFTSD-----YQSVKALLNG-KTFDPKSSEFTDFV--SKPFDLTNGRT 102
 Db 323 KASLANUTKEKERSAEKDSFERLNTALDINMEQENALFAGHLQOQYEFERKQES 382
 Qy 103 VLEIP---KKYQVVISFSPEDDKERPRGLFHLKEKLE-----DGNIAQSA TKFIYLLPL 154
 Db 383 LLKLETEHKQLQKRIGEFKIESEAKSEALLTQERELLEKREIDDLTLTQASLEYEQRR 442
 Qy 155 DMPKAAALQGYIYVDKNFNLI THPLSNFSAQSIKPLALTFSSDFIAKLNQFNQDELWV 214
 Db 443 NQ---VLKKEHRQVOQHFNQL-VHAKKK-----LDQKRYLAERKRI---DEEQI 485
 Qy 215 Y-----LEKPF-----DLEALKANIRLQTADFS---FEKGNLVDPF 247
 Db 486 FKLKEKIATERRELEKLYLVKKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 543
 Qy 248 V-----YSFIRNPQ---NQK-----EWA-----SDLN-----ODOKT 271
 Db 544 ASQKSLQKSFQKKNKEALNKAQIAEDWAHLKQNKHHADLEIFLEGFNFHLQOEKH 603
 Qy 272 VRLYRTEFSPQAKTILKYKDETFEL---SSIDLKASNGTSLFANENDLKQDLVDL 327
 Db 604 KLEARTQFQDNRVSLLSARFKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 662
 Qy 328 LDVSDYFGGQSETITTSNQKVPASERSLKDRVKFKDQKQKQKQKQKQKQKQKQKQKQ 387
 Db 663 EKQKEMLDGVHGFQFDENSLNISKLAEREL--AKFKEKELEAAQKQLSDNNNAGLKL 720
 Qy 388 QLQELVSKPNSIKDLVNATLARNRFLSLGKYNFLLFDDDLASHLDYFVLVSKAKIKQSSITK 447
 Db 721 QLDKLSLTKTERLELEASKERILDF-----YDESSRRRIADYSDQLARLAEVKTLE 772
 Qy 448 KLFIELPIKLSKSSILGDQEPNITKLFKEVTFKLDNFRD-----VRIEKAFLYLPGV 502
 Db 773 K-----NQETAAKS--ERELKVALEKLNQAKKAFQIRQQLLEIASV 814
 Qy 503 NEELEQAR---KAQRASFEK--EKSXKGLKFKFSQKQKQKQKQKQKQKQKQKQKQKQ 553
 Db 815 KQQLAQAKANLKNQQAELDKQTELEAAFLQDQTDKKELEKALHSVKSKQELLERERSFL 874
 Qy 554 ERLPENSPIQYQENAGIGASDPKPYMKVQVQORY-YLAKSQIQELIK-----601
 Db 875 -----LQKREFAEHVAGFKQVHFKTMMQRLSEFNKQKQKQKQKQKQKQKQKQKQKQ 926

```
QY 602 -AKDYTKLAKLNRHTYNTSLRKE-----QLFDVNPRIPSRDIERAKFVL 648
DQ 927 LKDY-QLFELQKQNFQEQKHEKLELLAQOAEKQELKQKATALASQDQDTQAKL 985
QY 649 DKTEKNYQWYISASVPFQKNSL-FGYRYLLGLDPKQTHIELVKLGKAGL----- 701
DQ 986 DLARQQHELELRQNAF-----NQASLSLQKQREQLTNQVKVYLHGELKKRHEKLTLDORLLA 1041
QY 702 -----QPEGYENLPSDFN-----LEDLKNIRIKTLPFSQKDNFKLSLLD 740
DQ 1042 EKEKQHKDAEINQRPQENYADFDQAKKELQELNQIR-----RNLEOSNASLLK 1095
QY 741 FNNYDGEIKAPFGPLFLPKELRNSSNGSSQNSPWEQEIISQFKD-QNLSNQ-- 797
DQ 1096 KRN-----QUTLDFAALLRKVQHTQNTQNRVQLNTQI---KEFLLEKKNFKQASDEAA 1143
QY 798 -----DQLAQFSTKIW---EKIIGDNEFDQNNRLOYLKLLDQESWINKTRDNLWT 847
DQ 1144 LQKALLIKLRSFASLQLOREALATQKLEFRD-----EQKSEINNKLQ----- 1192
QY 848 YLGDKLVKPKNNLEAKFRQISNLQELLTAFYTSAAALNNNNYQDSGAKSTIIFEEIAE 907
DQ 1193 ---EQPKLEKQNDQAKQKQLEFFQCCQRLDVEKLLKQKLVQKNLSKSYLYTKNRAD 1249
QY 908 LDPKVKKEGVADYVQLKFHYAIGFDDNAGKFNQEVIRSSRTIYLYKTSQKSLKLEADTIDQ 967
DQ 1250 LSQQQLQHKYANLELK-----EKLQAKRALDKKHRAIYQKMAQFVSE 1293
QY 968 LQAVKNAPIGLQSFLDTER-FGVQKLAISLAVQHKQKEKTLPKLNDGVTLLHDKL 1026
DQ 1294 LROEKKQL-LSAQKQVDDKSRLEQQRHLQNLQSLSETKKQKROGLEHDK-----FDOR 1346
QY 1027 KKPVIPOISSPEKQPEGLNQ-NGOSQNV--NVSTFGSIIESPSTNFQBDADLDD 1083
DQ 1347 RKAQVSSILNSHK-----KLKQEGELQILQKLSLKTQIEQ-FSKLYQKREKLDQ 1399
QY 1084 QDDSRQGNNSLQNEAGLKQKAILLGNQFIQYQONQKIEFEIINVERKVSLSFRV 1143
DQ 1400 RTTSLKHLRELKAQNEATAHKNREVLEIEN---YKKELQRLTTEKSEFDDNNKNLFEY 1455
QY 1144 EFKLAKTLEDNGKTIRVLSDTMSLIVNTTIETKPEMSAVPEVFDTKWVEQYDPRPLAA 1203
DQ 1456 FRIRNEIEKKEAHIKTVLEETQK-----KRHLVETEAVKHLHQKQSIIS 1500
QY 1204 KTKFVLKFKDQIPVDSGNSIDSKWLASIPLVHQQMLRLSPVVKTIRELGLKTEQOQOQ 1263
DQ 1501 KGOELKEIKERVSRDISH--TNKQREBELNSLHONKLLQKNLAEREREINNKSLLTKI 1558
QY 1264 QOQOQOQOQKKA-----VRKBE-----LETYN-----PKDEFNINLPLTKAHL 1303
DQ 1559 QTAQKLSKEARILKLEKQRAVEQOQYQAEITRLKTRNADLEKDNKHLFPPLFKIN-- 1616
QY 1304 TSLNVLNNDNY-----KIEDLVKINEAGDHLAFSLRANNTKRLMNTPTIFA 1352
DQ 1617 -----GNDMNYYPYFWFPOQKQSD-----SSNQIRHLFEQOQLQM 1653
QY 1353 DYNPFYFYNB--DWRSDTKYLNK-KNVSSHQOQAAGNQGSLIQLRLKNKI-----KPE 1404
DQ 1654 QQR---YENELTELRRQRLLEKLLQIQLESQLSAKKDNDFEKVEQMMQKLEKTEQKLS 1710
QY 1405 TFTPALIALKORNTNLSNYSVD 1426
DQ 1711 AFDQKINALAEQINTQKAEHAD 1732
```

RESULT 27

```
AAW20828
ID AAW20828 standard; protein; 2440 AA.
XX
AC AAW20828;
XX
DT 16-JUL-1997 (first entry)
```

```
XX
DE H. pylori cytoplasmic protein 11gel10309orf15.
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
OS Helicobacter pylori.
XX
PN WO9640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US009122.
XX
PR 07-JUN-1995; 95US-00487032.
PR 01-APR-1996; 96US-00630405.
XX
PA (ASTR ) ASTRA AB.
XX
PI Smith D, Berglindh OT, Mellgaard BL;
XX
DR WPI: 1997-052306/05.
DR N-PSDB; AAT68081.
XX
PT Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
PT useful for vaccines to treat or prevent H. pylori infection, and to
PT detect Helicobacter.
XX
PS Claim 61; Page 1233-1238; 1481pp; English.
XX
CC The present sequence shows a Helicobacter pylori cytoplasmic protein that
CC may be used in a vaccine to prevent or treat H. pylori infection or to
CC identify H. pylori polypeptide binding compounds, useful as potential H.
CC pylori life cycle activators or inhibitors. The genomic sequence of H.
CC pylori (ATCC 55679) was determined from overlapping contigs generated by
CC mechanically shearing the bacterial DNA. The sequences were analysed for
CC ORF of at least 180 nucleotides, and the predicted coding regions defined
CC by computer evaluation. To identify likely H. pylori antigens for vaccine
CC development, the amino acid sequences predicted from various ORF were
CC analysed for significant homology to other known or exported membrane
CC proteins. Having identified and determined the sequences of interest,
CC particular regions can be isolated from H. pylori by PCR amplification
CC for recombinant polypeptide production, e.g. in E. coli hosts
XX
SQ Sequence 2440 AA;
```

```
Query Match 2.9%; Score 280.5; DB 2; Length 2440;
Best Local Similarity 18.5%; Pred. No. 6,9e-07;
Matches 385; Conservative 316; Mismatches 678; Indels 697; Gaps 100;

QY 41 QGVISQLGLID----SVAFKPSI--ANFTSDYQSVKALLNGKTFDPKSEFTDFVSKFD 94
DQ 593 QG-LQQLGFNNDNHPKIFEPSLGTGKFTAHAPSCKNYRFIGTDLDPISANLSKFLYPNQ 651
QY 95 FUTNNGRTVLETPKKYQVVIS-----BF9PEDDK-----RPLGLHLEKLEDDG- 139
DQ 652 VIQNTALENYQFYQBYDAFVGPNPPYGNHKKIYSSNDKELSNESIHNYFLGKAIKELKDDGI 711
QY 140 -----NIAQSATKFIYLLPLDMPKA----- 159
DQ 712 GAFVSSWFMDAKNPMREHIAKNAT---FLGAILPNVSVFKATGAEVTSDIVFPKKGVE 768
QY 160 -----ALGOYSYIVDKNFNNL---IIHPLSNFSAQSIKPLALTRSSDFIAKLQNFN 207
DQ 769 KATNQSFTKAMPY---DKILNSLDDDTLFAQQNNRPSFIP-----SDQLKIVNAVA 818
QY 208 N-----QDELWYLEKFPDLBALKANIRLOTADFSP--EKGMLVDPFVYSFTRNPQOKE 260
DQ 819 NHFGFKQEKLRWYEK-----IDTANFGYSTQDYKIIKDFI-----DKV 857
QY 261 WASLDNODQKTVRLYLRTFESPOAKTI-----LKDYKQDETFLSSI---DLKASNGTSL 312
```

Db 858 GKSNINLEQTLNEY-----FIHPENILGHLSEKTYRRETNGEQIYKYDLOALEDESL 913
Qy 313 FANENDLKDOLDVLLDVSYFGQSETIITSNQVPPVPAASERSLKDRVFKDQKQPRI 372
Db 914 DLSQ-ALKQAEIKLPKDVQYQY-----HKTKTDLVDLIIDSNERYQEVQKLIKX-----L 962
Qy 373 EKFSLYEYDALSFYSLOE-----LVSKP-----NSIKOLVATL- 407
Db 963 ERRELVKWDLN--YFQLEQNNEMGIFLTKPKINSKVQDSRLKAYFKIKDALNDLTSALN 1020
Qy 408 ----ARMLRFSLGKYNFLFDDLAGHLDYFVLVSAKIKQSSITKKLFIELPIKLSKSSI 463
Db 1021 PLSSDLELENKRAKLNLYDEYVKKFY-----LNENKVK-----DIRQDL 1062
Qy 464 LGDOEPNKTLEKEVTFKLDNFRDVEIEKAFGLLYPGVNEELEQAKAQAASPEKEKSK 523
Db 1063 YGAKVLGLEKDFEKEITPRSAKMNIQPRQA-----QAKKAQ-IFPERTLNP 1108
Qy 524 KGLPEFSQOKEENSKAINQEGLE-----EDDNITERLPENSPIQOENAGL 571
Db 1109 KKELIITNAKEALIAJNQKGLDLHFIRDHFTTQSLTETIKELLEOKLIYKDHONGGY 1168
Qy 572 GASPD-----KPYMKQVONQRYVLAKSOIQ--ELIKAKDYTKLAKLLSNRHTYNIS 621
Db 1169 ILANDYLSGNVKKLKEVKEAINQVEGLEANVKDELILIPK----- 1211
Qy 622 LRLKEQLFDVN-PRIPSRDIEKAFVLDKTEKNKQIYSSASPVFQNKWSLFGYRYL 680
Db 1212 LKATEIMANINSPIPT---QYLEEFMLSELANSHYEKQYGDKMTDYQ----- 1255
Qy 681 LGLBPQTIHELVLGOKAGLOFGYENLPDSFNLEDLKNIRITPLFSQDNFKLSILD 740
Db 1256 ----LSNLKEDIKIEHLSG---AYEVFVRNNELNELYGRHDKPHSYKVPFESLNNK 1306
Qy 741 FNNYDGIKAPEFGLPLFLPKELRRNSNGSGQNSPWEQBIISQFQDNLSNQ--D 798
Db 1307 VLNNKOLSVKAOV-----DPNDPKKEIFITDEEQSNLARQKAE 1345
Qy 799 QLAQ--FSTKIWEKIIGDNEFDQNNRLOYLKLLKQLESWINKTRDNLVWYVLGDKLVKP 857
Db 1346 ELKEAFKDWIYK-----DYSRRTHLE-QIYNDTFNNVLKTYD-----GSQLELEG 1390
Qy 858 KN-----NLEAKPROISN---LOELLTAFYTSALSNNWYVDSGAKSTIIFEEI 905
Db 1391 FNYHISLRPHQKNALFRTIQDRAVCLDHQVAGKTLCAIASMEQKRMGLVKNKTLI---- 1446
Qy 906 AELDPKVEKGVADYVQLKFHYAIGFDDNAGKFOEIVRSSRTIYLTSGSKSLEADTI 965
Db 1447 -AVPNHLTKQWGEFYKAYPNANVLVDSKOTTEKE-----RELLFNQIANNYDAVVI 1499
Qy 966 DQLOAVKNAPLGOSFYLDTERFVGFVKLATSIAVQHKKQEKTLPKKLNDGYTLIHDK 1025
Db 1500 AHTHLELLSNPRGI-----IEELKEEELVNAEKNFERQELAYKN 1538
Qy 1026 LKKPVIPOISSPEKDWEGKLNQNGQSNVNVSTFGSIIESPSTFWQDADLDQDQ 1085
Db 1539 ----NPRETKPNBRAPKNDLK-----IRAKY-----DAILEKGS 1571
Qy 1086 --DDSRQGNLSLDQEGALK-----QKLAILLGNQ-----FI--QYVQNDK 1124
Db 1572 HIDIQMGIDNLIVDEAHLFKNLAFETSMEKIA-GLGNQOQSNRARDLFIKTRYLHNDK 1630
Qy 1125 EIEF----EIIINVKVSELSFVPEFKLAKTLEDNKGKTRVLSDETMSLIVNTTIEKTPM 1180
Db 1631 KIMFLGTGTPIAN--SISEMYHLQRYLTPDLVKERGL----- 1664
Qy 1181 SAVPEVDFTKWBOY-----DPRTPLAKT-KFVLKFKDQIPVDGSGNISDKWLASIPLV 1234
Db 1665 ----EFPD--DWAKTGEVNDPELDTSAQSYKVMNRFSKFSVQV---LSTMYRAFADIV 1716
Qy 1235 IHQOQMLRLSP-----VVKTIRE-----LGLKTEQOQOQOQOQ--OQOQOQPKK 1274
Db 1717 SNDDILKINPHFVKYVKGDKPINVVVVKSEEAQFIGVALENGKYNEGSIIDRMQKCEGK 1776

Qy 1275 AVRKEBELSTYNPKDFENILNPLTKAHLRLTSL-NLVNNDPNYKIEDLKVIKNEAGDQLA 1333
Db 1777 KSQKQGD-----NILSCTTDARKVALDYRLI--DPNAKVEK-----EFSKS 1815
Qy 1334 FSLRANNIKRLMNTPTIFADYNPFFYYNEDWRSIDKYLANKGNVSSHQOQOAGNQGSL 1393
Db 1816 YAMAKNIYENYLETHATKGTQLGFI-----GLSTPKTHSOKVS----- 1853
Qy 1394 IQRLNKNIIPETPTPALIAIKDRNN---TNLSNYSKIIIMIKPKYLVERSIGVPMSTGLD 1450
Db 1854 LEALDNAHETENKPL-----DKAELLESLSYDEKGNLIAP----- 1891
Qy 1451 GYIGSBOTKDTGSSSQKGFQDQFIQALGLKNTYHGKLGLSIRIFDPCNELAKIKDAS 1510
Db 1892 ----SKKELENELKEAKSVNLDIEAKGSDVDVLDVLRHLVQMGIPONEIAFIHDA- 1946
Qy 1511 NKKEBEKLLKSYDLFNKYLNEYEKK---SP-KIAKGWTNI-----HPDQ- 1550
Db 1947 --KTES---QKQDLFKK-LNRGGVRVLGSPAKMGVG-TNVQERLVAHHELDPCWRPDEL 1999
Qy 1551 ----KEYPNQKLPENYLNVL-----NOPW-----KVT 1576
Db 2000 LQMEGRGIROGNILHQNDPENFRMKIYRYATEKTYDSRMMQIIETKSKGIEQFRNAHKL 2059
Qy 1577 LYNSSDPITNLFVPEPSDRSGSTK-----LKQV-IQOVNNNYADWSAYLTFWY 1626
Db 2060 LNLBEP--NM-----GSSNASEKBAEATGNPLIIEVKLRABEIKSESKYKAFNKEHYF 2112
Qy 1627 DKNIITNOPNVIITANIADVFIDKVBLED-NTKLIAPNITO-WMPNISGSKEPY----- 1679
Db 2113 NEESLKN-----NASKLDYLKQLKOLETLQRSVITPTHTTEIKLYDLKNEBKDYELIKV 2167
Qy 1680 ---KPTVFFGNWENENSSMNSOQTPTWKIRE-----GFALOALKSSPDQKTRTF 1727
Db 2168 KEVEPL-----KENASMSSEL---THKKLKSQNKQIAEQNKKEKLDIAIKQFASNLT 2217
Qy 1728 VLITNAPLPLWKYGLGFI-----QNGPNFKTQDMRLVFQND----- 1763
Db 2218 FVNBEEDYKLEY--KGFVNNAVTKYQVEFSLSPKDNFNIAVSPSNVYKNDTINMFS 2275
Qy 1764 -----DNQIAALRVQBDPPEKSSE 1783
Db 2276 YNFCABIKFDGFLKRLDNATKL-----PEKIKE 2304
RESULT 28
ADE56037
ID ADE56037 standard; protein; 3259 AA.
XX AC ADE56037;
XX DT 29-JAN-2004 (first entry)
XX DE Human Protein CAA53052, SEQ ID NO 1876.
XX KW Human; pain; neuronal tissue; gene therapy;
XX KW spinal segmental nerve injury; chronic constriction injury; CCI;
XX OS spared nerve injury; SNI; Chung.
XX OS Homo sapiens.
XX PN WO2003016475-A2.
XX PD 27-FEB-2003.
XX PF 14-AUG-2002; 2002WO-US025765.
XX PR 14-AUG-2001; 2001US-0312147P.
XX PR 01-NOV-2001; 2001US-0346382P.
XX PR 26-NOV-2001; 2001US-0333347P.
XX PA (GEO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.
XX Woolf C, D'urso D, Befort K, Costigan M;
XX WPI: 2003-268312/26.
DR GENBANK; CA53052.
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
XX Claim 1; Page; 1017pp; English.
XX
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3259 AA;

Query Match 2.8%; Score 276; DB 7; Length 3259;
Best Local Similarity 18.8%; Pred. No. 2e-06;
Matches 310; Conservative 240; Mismatches 505; Indels 590; Gaps 76;

QY 106 IPKYQVWISFSPEDKFRFLGFLHKEKLEDPG-----NIAQSATKF 148
DB 1738 LSKKFSQSLMEKDSLSEEVQ-----DLKHQIDNVSKQANLEATEKHDNQTNTEGTQS 1792
QY 149 IYLLPLDMPKAAALGOVSYIVDRKNFNLIITHPLSNFSAQSIKPIALTRSSDFTAKLNQFNN 208
DB 1793 I-----PGETEEOQSL-----SMSTRPTCSSEVPSAKSANPAVSKD-----FSS 1831
QY 209 QDELWVYLEKFPDLEALKANIRLOTADFPSEKGNLVDPFVYFIRNPQNKQEWASDLNOD 268
DB 1832 HDEINYLQO---IDLKERI-----AGLEBEK-----QXNKGFSTLENE 1869
QY 269 QKTVRLVLRTEFSPQAKTILKYKYKDEFTPLSSIDLKASNGTSL-----FANENDLKQDLD 324
DB 1870 KNTLLSQISTK-DGELK-MLOEVTWKNLLNQIQEELSRVTKLKETAEERKDDLEERLM 1927
QY 325 VDILLVSDVPFGGSEFITS-----NSQKPV-----PASRSLSKORVKFKKQDQ 368
DB 1928 NQLAELNGSIGNYCQDVDAQIKNELLESEMKNLKKCVSELEBEKQQLVKEKTKVSEIR 1987
QY 369 KPRIEF--SLYEYDALSFYSQOELV-SKPNISKLVN-----ATLARNLR---- 412
DB 1988 KEYLEKIQAQKEPFGNKNHAKELQELLKEKQEBVQKQLOKDCIRYQEKISALERTKALEF 2047
QY 413 -----FSLGKNYFLFDLASHLDYFVFLSVKAKIKOSSI 445
DB 2048 VQTESQKDLTEITKENLAQAVEHRKKAQELASFKVLLDDTQSEA-ARVLDNLNKLKKELO 2106

QY 446 TKKLFIPIPIK---ISLKSSILGDOEPNITKTLFEKEVTFKLNFR--DVEIEKAFGLLYP 500
DB 2107 SNKESVKSMQKQDEDLERRLEQABEKHLKE--KKNQKQKLDALRREKVLHSETIGEIQV 2164
QY 501 GVN---EELEOARK-----AQRASPEK-----EKSXKGLKEFS----- 530
DB 2165 TLNKKDKVQOQOENLDSTVTQLAAFTKSMSSLDQDRDRVIDEAKKWERKFSDAIOSKEE 2224
QY 531 --QOKEENSKAINNO-----EGLEDDNITERLPENSPYQOENAGLGA 573
DB 2225 EIRLEKEDNCVLDKQLRQMSIHWEELKINISLEHDKQIWESKAO--TEVQLQKVCVDTLQ 2283
QY 574 SPDPKYMIKDVQYRYLAKSQIBELIKAKQVTKLAKLSNRHTYINISURLKEQLFDVNP 633
DB 2284 GENKE--LLSQLEBTRHLVHSSQ-----NELAKLESELKS-----LKQLTDL-- 2324
QY 634 RIPSSRDTEKAKFVLDKTEKNYQWQIYSSASPVFNQKWSLFGYRYVLLGLDPKQTIHEL 693
DB 2325 ---SNSLEKCK-----BQKGNLEGIIRQ----- 2344
QY 694 KLQKQAGLOFE--GYENLPSPDFNLE-----DLKNIRIKTFLPSQKDNFKLSLLD 740
DB 2345 ---QEADIQNSKFSYEQLETDLOASBELTSRLHEEINMKEQKIISLSGKEBAIQVAIAE 2401
QY 741 FNNYVDGEITKAPFGLPLPLPKELRRN-----SSNGSGSONSPWEQEIISOFKQONLSN 796
DB 2402 LRQHDKEIKE-----LENLLSQEENIVLBEENKKAVDKTNQ--LMETLTKIKKENIQO 2455
QY 797 QDOLAQFSTKI-----WEKIIGDENEPD-----QNNRL----- 824
DB 2456 KAQLDSFVKMSSSLQNDRRIVGDYQOUEERHLSILEKQDLIOEAAAENNNKKEIRGL 2515
QY 825 -----QYK-----LLKDLQESWI-----NKTDRNLDYWTYTLGD 851
DB 2516 RSHMDDLSENAKDAELIQYREDLNQVITIKDSQKQLLEVQLQONKELENKY-AKLEE 2574
QY 852 KUKVXPKN--LEAKF-----RQISNLOE----- 873
DB 2575 KLKSESEANEDLRRSFNALQBEKQDLSKEIHLKVSISOLTRQVTALEBEGTLGLYHAQL 2634
QY 874 -----LLTAFYTS-----NALSNNWNYODSGAKSTIIIFEEIAELDPKVEKGVADV 920
DB 2635 KYKEBEVHRLSALFSSSQKRIAELEBEELVCVQKEAAK-----KVGEIEDKLUKK----- 2682
QY 921 YOLK-FHYAIGPDDNAGFNQEVIRSSRTIYLTSGSKLEADTIDQLNQAVQNAPLGL 979
DB 2683 -ELKHLHDAGIMRNETETAERVAELARDIV-----EMEQKLL-WYTKENKGLTAQI 2733
QY 980 QSF-----YLDTERFGVFOKLATGLAVOHKQKEKTLPPKLNNDGYTLIHDKLUK 1027
DB 2734 QSFGRSMSSLONSRDHANEELDELKRYDAJGLKELAQLEQGL-----LNRE-----RDALL 2785
QY 1028 KPVIFQISSPPEKDFE-GKLNQNGSONVNVSTFGSIIESPY-----FS---TNFQEDA 1078
DB 2786 SETAFSMNSTEENSLSHLEKLNQQLLSKDEQLLHLSQLEDYSYNQVQSFSKAMASLQNER 2845
QY 1079 D-----LDQDQDQDS-----RQGNNSLDNQEBAGLKLAILLGNO 1114
DB 2846 DHLWNELEKFRKSEBGKQRSAAQSTSPAQVQSLKAMSSSLQNDRLDLKE-----LKNL 2900
QY 1115 FTQYYQON-----DKEIEFPIINVE-----K 1135
DB 2901 QOOYLIQINQEIHELPLKAQLOEQYQDKTKAFQIMQEBELRQENLSWOHELHURMEKSSWE 2960
QY 1136 VSELSFRVEFKLAKTLEDNG-----KTIIVLSDETMSLIVNTTIEKTPEMSAVPE-- 1185
DB 2961 IHERRMKOYLMAISDKQOQLSHLQNLRELASSSQTPKQVQVQROQASPTSASPDGS 3020
QY 1186 ---VFTDKWVEQYDPRTPLAAKTKFVLKPKF---DQIPVDGSGNISDKWLASIPLVIHQOM 1239
DB 3021 QNLVYETELL-----RTQNDLSLKEIHQKELRIQOLNSNFSQLLEBKNTLSIQLCDTSQS 3075

Qy 1240 LRLSPVVTIRELGLKTEQQ-----QQQQQQQQQQP-----OKKAV 1276
 Db 3076 LR-----ENQHYGDLNHCVALEKQVQELQAGPLNDVAPGAPQKNGV 3120
 Qy 1277 -RKEELETPYKDBFNILNPLTKAHLTLNLVNDPNVYKIEDLVKIKNEAGDHQL--- 1332
 Db 3121 HRKSDPELRFPQSFSAQ-----QQLCNT--RQEVNELRKLLEERDQVAAE 3168
 Qy 1333 -AFSLRANNIKRL-----MNTPI 1349
 Db 3169 NALSDVAERQIRRLHSEWDSRTPI 3193

RESULT 29
 ADE56033
 ID ADE56033 standard; protein; 3259 AA.

XX AC ADE56033;
 XX DT 29-JAN-2004 (first entry)
 XX Human Protein CAA53052, SEQ ID NO 1872.

XX Human; pain; neuronal tissue; gene therapy;
 KW spinal segmental nerve injury; chronic constriction injury; CCI;
 KW spared nerve injury; SNI; Chung.
 XX Homo sapiens.

XX WO2003016475-A2.
 XX 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.
 XX 14-AUG-2001; 2001US-0312147P.
 XX 01-NOV-2001; 2001US-0346382P.
 XX 26-NOV-2001; 2001US-0333347P.
 XX (GEO) GEN HOSPITAL CORP.
 XX (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;
 WPI; 2003-268312/26.
 XX GENBANK; CAA53052.

PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 101pp; English.

XX The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction

CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a human protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3259 AA;

Query Match 2.8%; Score 276; DB 7; Length 3259;
 Best Local Similarity 18.8%; Pred. No. 2e-06;
 Matches 310; Conservative 240; Mismatches 505; Indels 590; Gaps 76;

Qy 106 IPKKYQVWISFSPEDDKERFLGFLHKEKLEGG-----NIAQATK 148
 Db 1738 LSKKQSLMSEKSLSEVQ-----DLKHQIEDNVSKQANLEATEKHNDQNTVEGTQS 1792
 Qy 149 IYLLPLDMPKAAALQVSYIVDKNPNLIHPLNFSQAQSIKPLALTRSSDFIAKLNPNN 208
 Db 1793 I-----PGETEEDSL-----SMSTRPTCSSEVPSAKSANPAVSKD-----FSS 1831
 Qy 209 QDELWVLEKFFDLEALKANIRLOTADPFSEKGNLVDPFVYFIRNPQOKEWASDLNQD 268
 Db 1832 HDEINNYLQO---IDQLKERI---AGLEEEK-----QKNKFSOTLENE 1869
 Qy 269 QKTVRLVLRTEFSPQAQTIILKYKDETFELSSIDLKASNGTSL-----FANENDLKQOLD 324
 Db 1870 KNTLLSQISTK-DGELK-MIQSEVTKNLLNQIQEELSRVTKLKETAEBEKDDLEERLM 1927
 Qy 325 VDLDVDFYFGQSETITS-----NSQKPV-----PASERSLKDVRKFKKDDQ 368
 Db 1928 NQLAELNGSIGNYQDVTDAQIKNELLESEMKNLKKCVSELEEKQQLVKEKTVSEIR 1987
 Qy 369 KPRIEKF--SLYEVDALSFYSQLOELV-SKPNSTIKDLVN-----ATLARNLR---- 412
 Db 1988 KEYLEKIQGAQKEFGNGKSHAKELQELLKEQEQVKQKOCIRYQEKISALERTVKALEF 2047
 Qy 413 -----FSLGKNYFLFDLASHLDYVFLVSKAKIKOSS 445
 Db 2048 VQTESQKDLKITKENLAQAVEHRKKAQAEALASFVKVLLDDTQSEA-ARVLADNLKXKELQ 2106
 Qy 446 TKKLFIELPIK---ISLKSSILGQEPNITLPEKEVTFKLDNFR--DVEIEKAFGLLYP 500
 Db 2107 SNKESVKSQMKQKDEDLERRLEQAEKHLKE--KKNMQEKLDA LRREKVLHEETIGTQV 2164
 Qy 501 GVN---BELEQARK-----AQRASPEK-----EKSXKGLKEFS----- 530
 Db 2165 TLNKKDKVEQVQLOENLDSVTVTQLAAFTKSMBSLQDDRDRVIDEAKKWKERKFSDAIQSKEE 2224
 Qy 531 --QKEENSKAINNQ-----BGLEDDNITERLPENSPFIQYQENAGLGA 573
 Db 2225 EIRKEDNCSVTKQQLRQMSIHMEELKINIRLHEDKQIWESKQ--TEVQLQKVCDTLQ 2283
 Qy 574 SPDPKYMIDVQVQRYTLAKSQIOELIKADYTKLAKLLSNRHTYINISLRKEQLFDVNP 633
 Db 2284 GENKE-LLSQLEETRLHYHSSQ-----NELAKLESELKS-----LKDQLTDL-- 2324
 Qy 634 RIPSRRDIEKAFVLDKTERKNKYQIYSSAPVQNKWSLFGYRYLLGLDLPKQTIHEL 693
 Db 2325 ---SNSLECK-----EOKGNLEGIIRQ----- 2344
 Qy 694 KLGQKAGLOFE--GVENLPSPFNLE-----DLKNIRIKTLPFSQKDNFKSLLD 740
 Db 2345 ---QEADIQNSKFSYEQLEDTLOASRELTSRLHHEINKEQKIISLSGKEATQVATAE 2401
 Qy 741 FNNYDGIKAPKAPFGLPLFLPKELRN-----SSNSGGSQNSNSPWEQIISQFKDQNL 796
 Db 2402 LRQHQDKIKE-----LENLSQEESEENIVLEENKKAVDKTNQ--LMTLTKIKENIQ 2455
 Qy 797 QQDLAQOFTKI-----WEKIIIGDNEFD-----QNNRL----- 824
 Db 2456 KAQLDSFVKSMSSLDNRDRIVGDYQQLERHLSILEKQQLIQBAAENKKEIRGL 2515

Qy	825	-----QYK-----LLKQLQBSWI-----NKTRDNLVWTVLGD	851	FT	/note= "one of 21 complete segments of homology of 22 amino acid length"
Db	2516	RSMDLNSNAKLDABLIQVREDLNQVITIKDSQKQLLEVQLQNKLENY-AKLEE	2574	FT	494..515
Qy	852	KLKVKPNV--LEAKF-----RQISNLOE-----	873	FT	/label= tandem repeat
Db	2575	KLKESEANEDLRSFNALQEEKQDLSKTESIKVSIQUTROVTALQEEGTGLYHAQL	2634	FT	/note= "one of 21 complete segments of homology of 22 amino acid length"
Qy	874	-----LLTAFYTS-----AALSNNWNYQDSGAKSTIIFBEIAELDPKVKKEGVADV	920	FT	516..537
Db	2635	KVKEEVHRLSALFSSQKRIAELEELVCVQKEAK-----KVGEIEDKLK-----	2682	FT	/label= tandem repeat
Qy	921	YQLK-FHYATGFDNAGKFNQEVIRSSRTIYIKTSKSKLEADTTDQLNQAQVNNAPLGL	979	FT	/note= "one of 21 complete segments of homology of 22 amino acid length"
Db	2683	-ELKHLHDAGIMRNETETAERVAELARDLV-----EMEQLKL-MVTKENKGLTAQI	2733	FT	538..559
Qy	980	QSF-----YLDTERFGVFQKLATSLAVQHKQEKTIPLKLNNDGYTLIHDKLK	1027	FT	/label= tandem repeat
Db	2734	QSGFRSMSSLQNSRDHANEELDELKRYDASLKELAQKEQGL-----LDALL	2785	FT	/note= "one of 21 complete segments of homology of 22 amino acid length"
Qy	1028	KPVIQISSPEKDWFE-GKLNQNGOSQNVNVSFSGIIESPY-----FS-----TNFQEDA	1078	FT	560..581
Db	2786	SETAFSNMNSTEENSLHLEKLNQQLSKDEQLLHLSQLEDSYNQVSFSGKAWASLQNER	2845	FT	/label= tandem repeat
Qy	1079	D-----LDQGDQDS-----RQNNSLDNOEAGLLKQKLAILLGNQ	1114	FT	/note= "one of 21 complete segments of homology of 22 amino acid length"
Db	2846	DHLWNELEKPKRSEEGKQKRSAAQPSAPAEVQSLKKAMSSLQDRDRLLKE-----LKNL	2900	FT	582..603
Qy	1115	FIQYQON-----DKEIEFEIINVE-----K	1135	FT	/label= tandem repeat
Db	2901	QOQYVQINQIEITLHPLKALQEQYQDKTKAFIQMBELROENLSWQHHLHQLRMEKSSWE	2960	FT	/note= "one of 21 complete segments of homology of 22 amino acid length"
Qy	1136	VSELSFRVEFKLAKTLEDNG-----KTIRVLSDETMSLI VNTTIEKTPEMSAVPE--	1185	FT	626..647
Db	2961	IHERMKEQYLMATSDKQQLSHLQNLIRLSSSSQTQPLKVQYQKQASPESTSASPDGS	3020	FT	/label= tandem repeat
Qy	1186	----VFDTKWVEQDPTPLAAKTKFVLKFK---DQIPVDGSGNISDKWLASIPLVTHQOM	1239	FT	/note= "one of 21 complete segments of homology of 22 amino acid length"
Db	3021	QNLVYETELL-----RTQLNDSLKEIHKELRIQQLNSNFSQLEEKNTLSIQLCDTSQS	3075	FT	648..669
Qy	1240	LRLSPVVKTIRELGLKTEQQ-----QOQOQOQOQOQ-----OKKAV	1276	FT	/label= tandem repeat
Db	3076	LR-----ENQHVGDLLNHCAVLEKQVQELQAGPLNIDVAPGAPQKNGV	3120	FT	/note= "one of 21 complete segments of homology of 22 amino acid length"
Qy	1277	-RKEELETYPKDFENILNPLTKAHLRLTSLNVLNNDPNYKIBDLKVIKNEAGDHOL---	1332	FT	714..735
Db	3121	HRKSDPELRPQQSFEAQ-----QQLCNT--RQEVNELKRLLEERDQRVAE	3168	FT	/label= tandem repeat
Qy	1333	-AFSLRANNIKRL-----MNTPI	1349	FT	/note= "one of 21 complete segments of homology of 22 amino acid length"
Db	3169	NALSVABEQIRRLSEHSEWDSSTRPI	3193	FT	736..757
RESULT 30					
AAR46608					
ID	AAR46608 standard; protein; 1663 AA.				
XX	AAR46608;				
AC	AAR46608;				
XX					
DT	16-OCT-2003 (revised)				
DT	25-MAR-2003 (revised)				
DT	22-SEP-1994 (first entry)				
XX					
DE	Plasmodium falciparum erythrocyte membrane protein PfEMP3.				
XX					
KW	Plasmodium falciparum erythrocyte membrane protein; PfEMP3; malaria; antigen; epitope; vaccine; anti-idiotypic antibody.				
KW					
XX					
OS	Plasmodium falciparum; (Malayan Camp strain).				
XX					
Key	Location/Qualifiers				
FH	472..493				
FT	Region				
FT	/label= tandem repeat				

```
FT      amino acid length"
FT      890..911
FT      /label= tandem repeat
FT      /note= "one of 21 complete segments of homology of 22
FT      amino acid length"
FT      912..933
FT      /label= tandem repeat
FT      /note= "one of 21 complete segments of homology of 22
FT      amino acid length"
FT      934..946
FT      /label= partial_tandem_repeat
FT      949..967
FT      /label= tandem repeat
FT      /note= "one of 11 complete segments of homology of 19
FT      amino acid length"
FT      968..986
FT      /label= tandem repeat
FT      /note= "one of 11 complete segments of homology of 19
FT      amino acid length"
FT      987..1005
FT      /label= tandem repeat
FT      /note= "one of 11 complete segments of homology of 19
FT      amino acid length"
FT      1006..1024
FT      /label= tandem repeat
FT      /note= "one of 11 complete segments of homology of 19
FT      amino acid length"
FT      1025..1043
FT      /label= tandem repeat
FT      /note= "one of 11 complete segments of homology of 19
FT      amino acid length"
FT      1044..1062
FT      /label= tandem repeat
FT      /note= "one of 11 complete segments of homology of 19
FT      amino acid length"
FT      1063..1081
FT      /label= tandem repeat
FT      /note= "one of 11 complete segments of homology of 19
FT      amino acid length"
FT      1082..1100
FT      /label= tandem repeat
FT      /note= "one of 11 complete segments of homology of 19
FT      amino acid length"
FT      1101..1119
FT      /label= tandem repeat
FT      /note= "one of 11 complete segments of homology of 19
FT      amino acid length"
FT      1120..1138
FT      /label= tandem repeat
FT      /note= "one of 11 complete segments of homology of 19
FT      amino acid length"
FT      1139..1157
FT      /label= tandem repeat
FT      /note= "one of 11 complete segments of homology of 19
FT      amino acid length"
FT      1158..1173
FT      /label= partial_tandem_repeat
FT      /note= "partial segment of homology"
FT      1179..1193
FT      /label= tandem repeat
FT      /note= "one of 4 complete segments of homology of 15
FT      amino acid length"
FT      1194..1208
FT      /label= tandem repeat
FT      /note= "one of 4 complete segments of homology of 15
FT      amino acid length"
FT      1209..1223
FT      /label= tandem repeat
FT      /note= "one of 4 complete segments of homology of 15
FT      amino acid length"
FT      1224..1238
FT      /label= tandem repeat
FT      /note= "one of 4 complete segments of homology of 15
```

```
FT      amino acid length"
FT      1248..1260
FT      /label= tandem repeat
FT      /note= "one of 27 complete segments of homology of 13
FT      amino acid length"
FT      1261..1273
FT      /label= tandem repeat
FT      /note= "one of 27 complete segments of homology of 13
FT      amino acid length"
FT      1274..1286
FT      /label= tandem repeat
FT      /note= "one of 27 complete segments of homology of 13
FT      amino acid length"
FT      1287..1299
FT      /label= tandem repeat
FT      /note= "one of 27 complete segments of homology of 13
FT      amino acid length"
FT      1300..1312
FT      /label= tandem repeat
FT      /note= "one of 27 complete segments of homology of 13
FT      amino acid length"
FT      1313..1325
FT      /label= tandem repeat
FT      /note= "one of 27 complete segments of homology of 13
FT      amino acid length"
FT      1326..1338
FT      /label= tandem repeat
FT      /note= "one of 27 complete segments of homology of 13
FT      amino acid length"
FT      1339..1351
FT      /label= tandem repeat
FT      /note= "one of 27 complete segments of homology of 13
FT      amino acid length"
FT      1352..1364
FT      /label= tandem repeat
FT      /note= "one of 27 complete segments of homology of 13
```

```
Query Match      2.8%; Score 275.5; DB 2; Length 1663;
Best Local Similarity 20.5%; Pred. No. 7.7e-07;
Matches 332; Conservative 227; Mismatches 578; Indels 485; Gaps 82;

QY 74  LNK-----TDP-----KSSEFTDPSVKPFDLTNNGRVLEIPKYYQVISEFS 118
Db    ||| |||
QY 389 VNGKKLSTFPFKRRNKLKRRQQLHKPKQY-----KKYQKLE--- 431
QY 119 PEDDKERFRGLFHLKEKLE--DGNIAQSATKFIYLLPLDMPKALGOYSIVDKNFNNLI 176
Db    ||| |||
QY 432 -----REKRENDPGE-----PLNTPEIHVIRPSDMDKGENKSA 465
QY 177 IHPLSNFSAQSIKPLALTRSSDFIAKLNQFNQ--DELWVYLEKFFDLALKAIRLQTA 234
Db    ||| |||
QY 466 GHP---FKYQPTKGLKEYEES-HVSKDYQLHEPPTKLPY-ERGHVSREYQLDHEPPTK 520
QY 235 DFSFEKGNLVDPP-VYSFIRN--PONQK-----EWASLNQDQKTVRLYLRTFSFQAKT 286
Db    ||| |||
QY 521 LPEYKGVHSREYQLDNEVRDELPEYKGVHSREYQLD-NEGPSTLKEYDQTELAKGDI 579
QY 287 ILKDYKYDETFLLSSIDLKASNGTSLPANENDLKDQLD-----VDLLD 329
Db    ||| |||
QY 580 TNKPHEVSDEYDQTEL-----AKGKDITNKPHEVSDEYDQSELAKGKDTNKPHEVSDEYD 635
QY 330 VSDYFGGQSEITNSQVQVPASERSLKDRVKFKDQKPKRIEKFSLYEYDALSFYSOL 389
Db    ||| |||
QY 636 QTELAKGKEVTKNKHENLE-----EYNETDLAKGKEVTKNKH--ESVDEYDQ--SELAKG 686
QY 390 QELVSKPNSIKDLVNATLARNLRSLGKYNFLFDDDLASHLDYFVLVSKAKIKQSSITKLL 449
Db    ||| |||
QY 687 KDIITNKPHEVSDEYDOT-----ELAKGKEVTKNARENLEBEYNETDLAKGKE--VTNKA 737
QY 450 FIELPIKISLSSILGDQEPNITLFE-KEYTFK-----LDNFRDVEIEKAFGLLYPGVNE 504
Db    ||| |||
QY 738 RENL-----BEYNETDLAKGKEVTKNAHENLEBEYNETDLAKGKEVTKN-AHE 783
```


QY 239 EKHNLDPFVYSPFRNPONQKAWASDL-NQDQKTVRLYLRTFSPQAKTILKDYKYKDET 297
D 3328 DTINLLQP-----LANKESLKEARNRLESKINETVQTDGMTQOS-----VENTK----- 3371
QY 298 FLSSIDLKAGTSU---FANENDLKQO-----LDVLDLDSVDFGQOSETITSNSQ 346
D 3372 ---QAKIKAKQESSIAQTLINNGDASQDEVESTEIKLQKSELTN---SINHLTVNKE 3424
QY 347 VKPVPASERSLQKRVFKKQOQKPRIFKFSYLYEVDALSFYQLOQLVSKPSNISKDLVNAT 406
D 3425 ---PLETAKNQLQANI---DQKPFST-DGMTQOSVQSYERKLOEAKDKINSI---NNV 3471
QY 407 LARNLRFSLGKYNFLFDDLAHLDYVFLVSKAKIKQSSITKFLFELPIKISLASSILGD 466
D 3472 LANNPDVNAIRTN-----KVETEQINNEL-----TOAQGLTVD 3505
QY 467 QEP-NIKTLFEK-----EVTFKLDNFRDVEIEKAFGLLYPGVNEELEQAR-K 511
D 3506 KOPLINAKTALQOSLDNQPSTTGTEATINYNAKRQAEQ---VIQNAKIIENAQPS 3561
QY 512 AQRASFEKEKKGKLFQSO-----QKENSKAINNQOGLBEDDNTERLPEPNSPI-- 562
D 3562 VQOVSDEKSKVEQALSELNNAKSALRADKQELQOAYNQ---LIQPTDLNNKKPASITAYN 3618
QY 563 -QYQOENAGLGAS-----PDKPYMIKDQONQRYYLAKSOIQELIKAKDYTKLAKLLS 613
D 3619 QRYQOFSNELNSTKTNTDRILKEQNPADVNN-----ALNKVREV---QOKLNEARALLQ 3671
QY 614 NRHTYNISLRLEQKFDVNPRISS-----RDIEKAKFVLDDKTEKN 654
D 3672 NKEDNSALVRAKEQLOQAVDQVPSTEGMTQOTKDDYNSKQQAQOEISKAQOVDINGDAT 3731
QY 655 KYWQYTSASPVFONKWSLFGYRYLLGLDPKQTHLVLKGLQKAGLQFEGYENLPSPFN 714
D 3732 T---QOISNAKTNVERALEALNAKTKGLRADKEELQOAYNQLTQ---NIDTSGTKPASIR 3785
QY 715 LEDLKNIRIKTLPFQKONFKLSLDFNNYD-----GEIKA---PEGLPLFLPKELRR 766
D 3786 KYNEAKSRIQTOIDSAKNEANSILNDNPPQSVQVTAALNKIKAVQPELDKAIAMLNKEN 3845
QY 767 NSSNGSGS-----NSNSPWE---QEIIISQPKQNLNNOQLOAFSTKIWEKII--GDE 815
D 3846 KNALVQAKQLOQIIVNEVDPTQGMTDITANNYKSKREADEIQKA---QOILNNGDA 3900
QY 816 NE---PDQNNRLOYLKLLDQBSWINKTRDNLYWYVLGDKLVKPK-----NNLEAKFRQ 867
D 3901 TEQOITNETNRVQAI-----NAINKAKNDL---RADKSOLENAYNQLIQNVDTNGKK 3950
QY 868 ISNLQELLTAFYTSALSNMNYODSGAKSTIIFEEIAELDPKVEKVA---DVYQL 923
D 3951 PASIOOYQAA---ROAIETQYN-----NAKSE-AHQILENSNPSVNEVAQLOKVEAVQL 4001
QY 924 KPHYAIGFDNAGKFNQEVIRSSRTIYLTSGSKLEADTTIDQLN-----QAVK 973
D 4002 KUNDALHILQON---KENNSALVTAKQLOQSVNDQPLTTGTMTQDSINNYEAKRENAQSAIR 4059
QY 974 NAPLGLQ-----SFYLDTERFGVQKLATSLAVQHHKQKEKTLPKKLNDGYTLIHDKLKK 1028
D 4060 NAEAVINNGDATAKQISDEKSKVEQALA-----HLNDAKQOULTADTTTEL 4103
QY 1029 PVIQPISSPEKDFEGKLNQNGQSONV---NVSTFGSIIESPSTNPFQEDADLDQDQ 1085
D 4104 QTAVO-----OLNRRGDTNNKPKRSINAYNKAIOSL----- 4134
QY 1086 DDSRQGNNSLDNQEAGLKQKLAAILGNQFIOVYQONDKEIFEIINVEKVS-----L 1139
D 4135 ---EQUTSADKANAVIQPIRTVQEVNNALQOVNQLNOQLTEAINQLQPLSNNDAKAA 4192
QY 1140 SPRVEFKLAKTEDNGKTRIVLSDETMSLIVNTTIETKTEPMSAVPEVFTKWQVDPRT 1199
D 4193 RLNLNENKINQTVQDGMT-----QOSIEAYQNAKRAVQNESNT 4230
QY 1200 PLAAKTKFVLKFKDQIPVDGSGNISDKWLASIPLVIHQOMLRLSPVVK---TIRELGLKTE 1257

Db 4231 ALA-----LINNGDADEQOITETDRVNOQTNTLTOAINGLTVNKEPLET- 4275
QY 1258 QOQOQOQOQOQOQOQKAVKKEELETYNPKDFENILNPLTKAHLRLTSLNANNDPNYKI 1317
D 4276 -AKTALQNNIDQVPSDGM-TQOSVANYNOKLO-----IAKNEINTINNVLANNP----- 4323
QY 1318 EDLKVIK-NEAGDHOLAFSL-RANN-----IKRLMNTPI---TFAD---YNPFF 1358
D 4324 -DVNAIKTKABAEERTISNDLTQAKNNLOVDTPQLEKIKRQLODEIDOGTNTDGMTQDSVD 4382
QY 1359 YYNEDMRS-----IDKYLNNKGNVSSHQOQAGGQSGSLIQ----- 1395
D 4383 NYNDSLSAAIIIEKGKVNKLKRNPTVEQVKESVANAQO---VIQLOQVARTSLVPDKTQL 4439
QY 1396 -----RLANKWIKPETPTPALIALKDRNNTNLSNYSKDIIMIKPKYLVERSIGVPWSGLD 1450
D 4440 QEAKNRLENSINOQTTDGM-----TQDSLNNYNDK--LAKARQNLKELI-----SKVLG 4486
QY 1451 GYIGSEOTKDTSSSQKGFDDQDQFIOALGLK-----NTEYH-----GK 1489
D 4487 QOPTVASIRQNTDEANAHK-----QALDTARSQTLNREPIYNHINSHLNAQKON 4539
QY 1490 LGLSIRIFDPCNELAKIK-----DASNNKGBEKL-----KSYD- 1523
D 4540 FKAQVNSAPNHTLETIKNKADTLNQSWMTALSESIADEVQKQOENYLDASNNKRQDYDN 4599
QY 1524 ---LFPKYLNE-----YEKSPKIAKGNTHIHDPQKEYPNPNQKLPENYLVNLV-- 1568
D 4600 AVNAAKGILNQTSPTMSADVIDQAEVDVKRTKTALDGNORLEVAQOAL--NHLNTLND 4657
QY 1569 LNPQWKVTLNNSDFITNLFVEPEGSDRGSGTKLKQVIQOVNNYAD--WGSAYLTFWY 1626
D 4658 LNDAQRTLDTINHSPNINSVQAKKANTVNTAMTQKOTIANYDDELHDGYNINADK 4717
QY 1627 DK-----NIITQPNVITANIADVIFIKDVLEDNTKLIAPNIITQMPNIPSGSKFKYKP 1681
D 4718 DKDAYNNAVNAQNLINQSDANQALDPAEINKVQRTVNTTKDNLGNDKLAERKRDAN 4777
QY 1682 TVFFG-----NWNEN-----SSNNSQAOPTWEKIREGALQALKSSFD-----Q 1722
D 4778 TTIDGLTYLNEAQRNKAENGVKASTKNTIITSOLOYNQL--NTAMQALRNSVNDVNNVK 4835
QY 1723 KTRTFVLTTNAPLWKYGP-----LGFQNGP-----NFKTQDMLVFNQDNDNQIAA 1769
D 4836 ANSVIINEDNGPKAYNQAVTHAQTLINAQBNPEMSRSDVNVNKTQAVNTAHQNLHGQOKL 4895
QY 1770 LRVOEQ-----DRPEKSSDKDKQKWKIKFKVPIPEMFNSGNIR 1808
D 4896 EQAQSSANTEIGNPLNTNTQKAKEK-----ELVNSKQTR 4930
RESULT 32
AD084851
ID AD084851 standard; protein; 10203 AA.
XX AC AD084851;
XX DT 29-JUL-2004 (first entry)
XX S epidermidis surface anchored LPXTG prbtein SeqID23.
XX LPXTG; cell wall-anchored surface protein; Gram positive bacterium;
KW extracellular matrix molecule; sequence database; C-terminal;
KW immunoglobulin-like fold region; Ig-like fold region; antibacterial;
KW vaccine; gene therapy; infection; medical device; prosthesis;
KW premature newborn; AIDS; debilitated cancer; bone marrow transplantation.
XX Staphylococcus epidermidis.
OS
XX WO2004025416-A2.
PN
XX 25-MAR-2004.
PD


```

Qy 512 AQRASFKEKSKGGLKEFSQ-----QKENSKAINNQGLEBDDNITRLEPENSPI-- 562
Dy 3562 VQOVSDEKSKVEQALSELNNAKSALRADKBOLOQAYNQ-----LIQPTDNNKKPASITAYN 3618
Qy 563 -OYQOENAGLAS-----PDKPYMIKOVQNYRYLAKSQLOELIKADYTKLAKLLS 613
Db 3619 QRYQOQFNELMSTTNTDRILKEQNPVSADVNN-----ALNKREV--QOKLNEARALLQ 3671
Qy 614 NRHTYNTISLRLEKDFVNPRISS-----RDIEKAKFVLVDKTEKN 654
Db 3672 NKEDNSALVRAKEOLOQAVDQVPSTEGMTQOTKDDYNSKQQAQOELSKAQOVIDNGDAT 3731
Qy 655 KYWQIYSSASPVFNKWSLFGYRYLLGLDPKQTHIELVLGQKAGLOFGYENLPSDFN 714
Db 3732 T--QQISNAKTNVERALEALNNAKTGLRADKEELQNAYNLTQ-----NIDTSGKTPASIR 3785
Qy 715 LEDLKNIRIKTPLFSQDNPKLSLLDFNYYD-----GEIKA--PFEGLPFLPKELRR 766
Db 3786 KYNEAKSRIQOIDSAKNEANSILITNDPNQVSVTAALNKIKAVQPELDAKIAMLKNKEN 3845
Qy 767 NSSNSGGSQ-----NSNSPWE---QBIISQFQDNLSNDOLAQFSTKIWEKII--GDE 815
Db 3846 NNALVQAKQLOQOIVNEVDPTQGTWTTDANNYKSKKEAEDEIQA-----QQIINNGDA 3900
Qy 816 NE---FQNNRLOYLKLKDLQESWINKTRNLWYTLGDKLVKVPK-----NNLEAKFRQ 867
Db 3901 TEQOITNETNRVNOAI-----NAINAKNDL-----RADKSQLENAYNQLIQNVDTNGKK 3950
Qy 868 ISNLQELLTAFYTTSAALSNNWYQDSGAKSTIIFEIEIAELDPKVKEKVGA-----DYOQL 923
Db 3951 PASIQOYQAA---RQAIETQYN-----NAKSE-AHQILENSPNWSNEVAQLOKVEAVOL 4001
Qy 924 KFHYAIGFDDNAGFNQEVIRSSRTYLYTKTSGKSKLEADTIDOLN-----QAVK 973
Db 4002 KYNDAIHLQN--KENNSALVTAKNQLOQSVDQPLTGTHTQDSINNYEAKRNEAQSAIR 4059
Qy 974 NAPLGLQ-----SFYLPDTERFGVFQKLATS LAVHQKQKETLPKPLNNDGYTLIHDKLKK 1028
Db 4060 NAEAVINNGDATAQAQISDEKSKVEQALA-----HLNDAKQOLTADTTTEL 4103
Qy 1029 PVIQISSSPKDFWFKGLNQGOSQNV---NVSTFGSIIIESPYFSTNFQEDADLDQDQG 1085
Db 4104 QTAVQ-----QLNRRGDTNNKPRISINAYNKAIQSL----- 4134
Qy 1086 DSRQGNNSLDNOBAGLLKOKLALLGNQFOITYQQNDKEIEPEIINVEKVSE-----L 1139
Db 4135 --ETQITSKDNANAVTQKPIRTVQEVNNALQVNLQNLQOULTEAINOLOPLSNNDALKAA 4192
Qy 1140 SFRVEFKLAKTLEDNGKTIIRVLSDETMSLIVNTTIEKTPEMSAPVEFDTKWVEQYDPR 1199
Db 4193 RLNLENKINQTVQDGMT-----QOSIEAYONAKRVAQNESNT 4230
Qy 1200 PLAAKTFVLKFKDOI PVDSGNSIDKWLASIPLVIHQOMLRLSPVVK--TIRELGLKTE 1257
Db 4231 ALA-----LINNGDAEQOITETDRVNOQOITNLTAINGTUNVAKPELET- 4275
Qy 1258 QOQOQOQOQOQOQOKKAVRKEBELETYNPKDEFNLPLTKAHLRLTSLNLVNDPNYKI 1317
Db 4276 -AKTALQNNDQVQSTQGM--TQQSVANYNQKLQ-----TAKNEINTINNVLANNP----- 4323
Qy 1318 EDLKVIK-NEAGDHLQAFSL-RANN-----IKRLMNTPI---TFAD---YNPFF 1358
Db 4324 -DVNAIKTNKAEERIISNDLTQAKNNIQVDTQPLEKIKRQLODEIQGTWTDGWTQDSVD 4382
Qy 1359 YYNEDWRS-----IDKYLNNKGVSSHQOQAAGNGQSGGLIQ----- 1395
Db 4383 NYNDSLSAAIIEKGKVNKLKRNPVTPEQVKESVANAOQ---VIQDLQNARTSLVPDKTQL 4439
Qy 1396 ----RLNKNIKPETFETPALIALKDRNNTNLSNYSDAKIIMKPKYLVERSIGVPWSGTGLD 1450
Db 4440 QEAKNRLENSINQOTDTDGM-----TQDSLNNYNDK--LAKARQNLEKI-----SKVLG 4486
Qy 1451 GYIGSEOTKGTSSSSQKQFDQDIFQALGLK-----NTSYH-----GK 1489

```

Db	4487	GOPTVAEIRQNTDEANAHK-----QALDTARSQJTLNREPYNINHNESHLLNNAQKON	4539
Qy	1490	LGLSIRIFDPGNEELAKIK-----DASNKKEEKL-----KSYD-	1523
Db	4540	FKAQVNSAPNHNLTETIKKADTLNOSMTALSESIAIADYENKQOENYLDASNKKRQDYDN	4599
Qy	1524	---LFRNYLNE-----YEKSPKIAKGWNTNIHPDQKEYPNPNQKLPENYLNLV--	1568
Db	4600	AVNAAKGILLNQTSPTMSADVIDQKAEDVKRTKTALDGHQRLBEVAKQAL--NHLNTLND	4657
Qy	1569	LNQPKWVTLYNSSDFITNLFVEPEGSDRGSGTKLKQVIOQVNNYAD--WGSAYLTFWY	1626
Db	4658	LNDAAQRQTLTDTIHNSPNINSVNOAKEKANTVMTATQLKQTTIANYYDDELHDGNYINADK	4717
Qy	1627	DK-----NIITNQPNVITIANADVFIKDVELEDNFKLIAPNITQWPNISGSKKFKYPK	1681
Db	4718	DKKDAYNNAVNAKQLINOSDANQAQLDPAEINKVTVQVNTTKDNLGNDKLAEKRDAN	4777
Qy	1682	TVPFQ-----NWNEN-----SSMNSQAQTPTWKIREGFALQALKSSPD-----Q	1722
Db	4778	TTIDGLTYLNEAQRNKAENVGKASTKTNITSQLQDYNQL--NIAMQALRNSVNDVNNVK	4835
Qy	1723	KTRTFVLTNAPLPLWKYGP-----LGFQNGP-----NPKTDWRLVFQNDNDQIAA	1769
Db	4836	ANSNYINEDGPKPEAYNQAVTHAQTLLINAQSNPEMSRDVVVNQKTOAVNTAHQHLHQQKL	4895
Qy	1770	LRVQEQ-----DRPEKSESDKQKWKFKVVIPEPMFNSGNIR	1808
Db	4896	EQAQSSANTEIGNPLNLTWQKAEK-----ELVNSKQTR	4930
RESULT 34			
AAB18287			
ID	AAB18287	standard; protein; 1997 AA.	
XX	AAB18287;		
XX	07-NOV-2000	(first entry)	
XX	Plasmodium falciparum	chromosome 2 related protein SEQ ID NO:145.	
XX	Plasmodium falciparum;	chromosome 2; human malaria parasite; vaccine;	
XX	antimalarial;	malaria; protozoacide; infection; insecticide.	
XX	Plasmodium falciparum.		
XX	WO200025728-A2.		
XX	11-MAY-2000.		
XX	05-NOV-1999;	99WO-US026796.	
XX	05-NOV-1998;	98US-0107131P.	
XX	(HOFF/) HOFFMAN S.		
XX	(CARU/) CARUCCI D.		
XX	(GARD/) GARDNER M.		
XX	(VENT/) VENTER J C.		
XX	Hoffman S, Carucci D, Gardner M, Venter JC;		
XX	WPI; 2000-365347/31.		
XX	Proteins encoded by chromosome 2 of the human malarial parasite,		
XX	Plasmodium falciparum, useful as antimalarial vaccines and in the		
XX	diagnosis of P.falciparum infection.		
XX	Disclosure; Page 341-346; 577pp; English.		
XX	The present invention describes proteins and their fragments (I) encoded		
XX	by chromosome 2 of the human malarial parasite, Plasmodium falciparum.		
XX	Also described are: (i) nucleotide sequences (ii) encoding (I); and (2)		


```
QY 898 STIIFEEIAPLDPKVKEKVGAD-----VYQLKPHYAIGFDD---NAGKFNQEVIR 944
D 961 ASDLDKVKIEITDTLHAELQAERSSSSALHKLKSFDEIATGHKELTSKADAWSQEMLQ 1020
QY 945 -----SSRTIYLTSG--KSKLEADTIDOLNQAVKNA-----PLGLQS 981
D 1021 KEKELQELRQQLQSDQSDTKLKAEBERKEKSPESIKNLQEEVTKAKTENLELSTGTQ 1080
QY 982 FYLD-TERFGVFKLATSIAVQHKQKEKTLPKKLNNDGVTLIHDKLKPVPQISSLSPK 1040
D 1081 TIKDLQERLEI-----TNAELQHKWASEDAQIADLKLVE-----ATQVA----- 1123
QY 1041 DMFEGKLNQGNQSNVSTFGSIISPYFSTN-----FOEDADLDQD-----QDD 1087
D 1124 -----NANISATNAELSTVLEVLQAEKSETTHIFELFEMEADMNSERLIEKVTKIKEE 1176
QY 1088 SROGNNSLNQEAGL--LKQKLAILLGNQFIQYQNDKEIEPEI--INVEKVSLSFRVE 1144
D 1177 LKETHIQDLDERQKKFELBEKL-----KQAQOSEKLOQESQTSKEKLETIQ----- 1223
QY 1145 FKLAKTLENGKTIKIRVLSDETSIIVNTTIEKTPMSAVPEVDTKWVQYDPRTPLAAK 1204
D 1224 -QSLQELQDSVKQ-----KEE-----LVQNLEEKVRESSIIIEAQTKLNES---NVQLENK 1271
QY 1205 TKPVLKPKDQIPVDGSGNISDKWLASIPLVIHQOMLRLSPVVKTIHRELGLKTEQOQOQO 1264
D 1272 TSC-----LKETQDQLLES 1285
QY 1265 QOQOQOPKKAVERKEBELETYNPKDFEINILNPLTKAHLRLTSLNVLNNDPNYKIEDL-KVI 1323
D 1286 QKKEKLOQBEAKLSGELQ-----QVQEAAGDITKDSLKVVEELVRLV 1327
QY 1324 --KNEAGDHQALFSLRANNIKRLMNTPIFDADYNPFYINEDWRSIDKYLNNKNGVSS-- 1379
D 1328 EEKLOATSQLDAQATN--KELOELLV-----KQSENEGNIQGES 1366
QY 1380 -----HQQAAGNGSGSLQIRLN--KNIKPETTPALIAKDRNNTNL-----SN 1423
D 1367 LAVTEKLOLQEQANGELKEALCKQENGLKELQGL-----DESNTVLESQKKSHNE 1417
QY 1424 YSKTIIMIKPKYLVERSIGVPWSTGLDYGIGSOTKDTGTSSSSQOKGFQDFOALGLKN 1483
D 1418 IQKLSEQAQOK---ERTLOEETSKEALQESQLKQANEELQKSLQOK-----QLLEKG 1467
QY 1484 TEYHGKLGSLIRIFDPGNELAKIKDA 1509
D 1468 NEFDTQLAEYQKVIDEMDDAASVSKA 1493

RESULT 37
AAB18195
ID AAB18195 standard; protein; 1516 AA.
AC AAB18195;
XX
XX
XX 07-NOV-2000 (first entry)
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:52.
DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
XX Plasmodium falciparum.
OS
XX
XX WO200025728-A2.
XX
XX 11-MAY-2000.
XX
XX
XX 05-NOV-1999; 99WO-US026796.
XX
XX 05-NOV-1998; 98US-0107131P.
XX
XX
```

(HOFF/) HOFFMAN S.
(CARU/) CARUCCI D.
(GARD/) GARDNER M.
(VENT/) VENTER J C.

Hoffman S, Carucci D, Gardner M, Venter JC;
WPI; 2000-365347/31.

Proteins encoded by chromosome 2 of the human malarial parasite,
Plasmodium falciparum, useful as antimalarial vaccines and in the
diagnosis of P.falciparum infection.

Disclosure; Page 120-124; 577pp; English.

The present invention describes proteins and their fragments (I) encoded
by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
vaccines against P. falciparum infection comprising (I) or (II). (I) and
(II) are useful for the development of vaccines against P. falciparum
infection. (I) and polyclonal antisera or a monoclonal antibody raised to
immunogens comprising the sequences of (I), are useful in the detection
of infection with P. falciparum. Furthermore, (I) especially when they
are rifins or secreted or membrane proteins) can aid the identification
of drugs to treat or prevent P. falciparum infection, or they can be used
to identify drug resistance in P. falciparum. Sequencing of the
Plasmodium chromosome 2 and the subsequent identification of proteins
encoded by it will help to expand our understanding of parasite biology,
a process hampered by the complexity of the parasitic lifecycle, and
provide new targets for vaccine and drug development. Parasite resistance
to drugs and mosquito resistance to insecticides have led to a resurgence
of malaria in many parts of the world, and there is a pressing need for
vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
represent nucleotide and protein sequences given in the present
invention, but which are not specifically mentioned within the
specification

Sequence 1516 AA;

Query Match 2.8%; Score 272; DB 3; Length 1516;
Best Local Similarity 18.6%; Pred. No. 1.1e-06;
Matches 317; Conservative 254; Mismatches 543; Indels 594; Gaps 82;

QY 36 GVNPTQGVISQGLI-----DSVAFKPQITANTSYQSVKALLNGKTFDPKS----- 83

D 2 GVKGLWSIVSPVGVRVNPEIFTGKRIADVGIWLYELTYANNVKOLRN-KSFDNMSIFND 60

QY 84 --SEFTDFVSKFDFLTNNGRTV-----LEIPKKY-----QVVISSESPEDDKERFRLGF 130

D 61 LWIDFSENISS-EIKTDNIKKAHLYFFFLRICKLLYNIRPIFDGPNPELKRKTIFOR 119

QY 131 HLKKELEDGNTAQSAKFIY-----FNNQDELWVYLEKFDLEALKANIRLQATDFSEKGNLV 150

D 120 NIKKRYNEBEKPKTKAEKLVYNYQRTLLNSMKSKNKDNNSNNIEDKTNTPKNTQNS 179

QY 151 -----LLPLDMPKALQO--YSYI--VDKNFNLLIHLPLSNFSAQSIKPLALTRSS 197

D 180 NTQNSKNTPNKINADISKSLIQIYDDIKKDKSLNSLVEH-VGNVPV-SVKDVL-LTICN 236

QY 198 DFIAXLQ-----FNNQDELWVYLEKFDLEALKANIRLQATDFSEKGNLV 244

D 237 DDLISKIKNIKPIMTDFGPVLFGEQDGMGTVENTINKLD-----NRNKDNNULS 285

QY 245 DPFVYSFIRNPQNKQEWASDLNQDQKTVRLYLRTEFSPQAKTILKDYKYKDTFLSSI-- 302

D 286 YSINYNKVQDVNNND--DDKDKKENINEVRDQ-----KNYVYKKNENINNYL 334

QY 303 -----DLKASNGTSLFANENDLKDQ-----DVDLL 328

D 335 DDDDEKEDIQNKGV---YNNDDIDEQIRKHKHMARKKYIESIPKTFKGLCWRPRVDII 391

QY 329 DVSDYFGGQSETITSNQKVPVPSERSLKDVRKFKDQOKPRIEKFSLYEYDALSFYSQ 388

Db	392	DISNY-----NTEMLEI---SETLKVHKNFK-----	415
Qy	389	LQEL-VSKPNSIKOLVNATLARNLRFSLGKYNFLDGLASHLDYYPVLSKAKIKQSSITK	447
Db	416	-QHLAVLDENNSTPVMNLLKNINY--KKN--DDLIEGGE-----K	452
Qy	448	KLFLFELPKISLSKSIILGDQEPNITKLFKEVTFKLDNFRDVELEKAFGLLYPGVNEELS	507
Db	453	KSFINL---INVD-----CYSSNSRLE	473
Qy	508	QARQAQASPE---KEKSKGLKEFSQOKEENSAINNOEGLEEDDNITERLPENSPIQ	563
Db	474	NDENIERGKINMFTNDEKSNININNNNNNNNN--NNDNMDNDVILIEHKNKNMIY	532
Qy	564	YQENAGLGASDPKPYMIKOVORYYLAQSOIELKAKDYTKLAKLLNHRHTYNSLR	623
Db	533	-----DNKYNVE-----CSSEKINDNGISNKNINILEPNNLDTSNIFLE	572
Qy	624	LKEQ--LPDVNP---RIPSSRDIEKAKFVLDKTEKNKYWOIYSSASPVFQNKWSLFGYYR	678
Db	573	GKDEYKYVYNKBEIRIPLFKEINK--IFEKPLKLYQILQD-----IKEWTNDRIK	626
Qy	679	YLLGLDPQTHLELVKLQKAGLQFEGY-ENLPSDFNLEDLK-----NIR-IKTPLFSQK	731
Db	627	AIKSKDDMVFSQ-----VQLETVMRIKTDFFEIKLKIKMAENIQSVGELLINK	677
Qy	732	DNFK-----LSLLDFN-----NYDGEIKAPFGLPL- 758	
Db	678	DLKNTDINIKDYNLVGLKKKKKKKFLNDILNTYNTFTTESKYQDLYVKGESSKEDIK	737
Qy	759	-----FLPKELRRNS-----SNSGGSQNSPWEQELISQFKDQLSN	796
Db	738	NQIDFVTQECYRNNDIIRDTKSDIFKNYIKIDNNKYEIYNLELEQEEINEKKNKNKN	797
Qy	797	QDLAQFSTKWEKIIGDENEF-----DONNRLOYKLLDLOESWINKTRDNL-----	844
Db	798	NDSNKTFFLKI-----ENEFKDLLDDSQIFGDSLLADIKE--YNYTADNLDNNEN	848
Qy	845	-----YWT-----YLG-----DKLKVKPKNNLEAK-- 864	
Db	849	KSLYEDGENFITRNEPITNEYEKNIIYISDEQKYNEEDIIFKDKIKEKEKNNDTSDD	908
Qy	865	FRQISNLQELLTAFTYSAALSNMNYQD-SGAKSTIIFEEIAELDPKVEKGVADYVOL	923
Db	909	FENC-S-VOEKI---YVNEKIEBYNNKNDKSSSSSIILEELIKYKKEKDSLSPNLCVL	964
Qy	924	--KFHYAIGFDNAGKFNQEVIRSSRTIYLTSGSKLEADTIIDQLNAQVKNAPLGLQS	981
Db	965	LDEFHSNDLENN-----YISVSSD-----DMKTNVSKNITGVKE	1000
Qy	982	FYLDTERFQVFOK-----LATSLAVOHKQKKTLPKKNL-----NDGYTLIHDKLKPVIPO	1033
Db	1001	NKVDTNVEYDKGDDGVITISFEDSHKLESKFDNNNIYDND-----DELEKNLSKD	1054
Qy	1034	ISSSEKDFWFEKLN-QNGOSQNVNVSTFGSIIESPYSTNPFQEDADLDQDQDSDRQGN	1092
Db	1055	YISDVDKHVNNIYNIERGEREN-----EFVENKIQST-----ESHKSN	1095
Qy	1093	NSLNDQEAELLKOKLA-----ILGNOFIOYQONDKEIEFEIINVEKVSLSFRVE	1144
Db	1096	EFICTENKSLRQYWSKEDISNVRILKSDDDINLSKQN-----YFELLDDKKQVMDFQWN	1151
Qy	1145	F-----KLARTLEDNGKTIRVLSDETMSLIVNTTIKTPEMSAVPEVFTKWEQYDERT	1199
Db	1152	IEQNNDKLAKEDKLDEGAFYEYLED--NKIIDSYIKETNKE-----NEELIKEY----	1197
Qy	1200	PLAAKTKFVLKFKQDQIPVDGSGNISDKWLASI-----PLVHQOMLRLS-----PV	1245
Db	1198	-----KKLK-----KNINTEINDEMNDIDKLLNFFGPIYIQSPCEAQAQCSLYLNKNKYCAI	1249
Qy	1246	VKTIRELGL---KTEOQQQQQQQQQQQPKKAVRKEELETYNPKDEFNI-----LNPL	1297
Db	1250	ISDSDVLVFGSKGTIVKFNFNKKTVVEYKKAII--BEKGLGLYQ-EELINISLLCGDYT	1306
Qy	1298	TKAHLRLTSLN---LVNNDPNYKIEDLKVIKNEAGD--HOLAFSLRANNIKRLMNT-----	1347
Db	1307	IGVHGIGIVNALEIIFKAFNF--EDLKLDIVSNPFRKIDKNMYNEEQOFLNTHKNYK	1364
Qy	1348	-----PITFADYNPFFYVYNEDMRSIDKYLNNKNGVSSHQOQAAG	1387
Db	1365	LNWIFPNFPDREVYKCFKVPKVCYTDIKKFEWHVPDITKTKFPLHKTWTNISEKVL----	1420
Qy	1388	NQSGGLIQRNLKNI-----KPETFTPAL	1410
Db	1421	NVLNPLIKYQNVNVRTYQSKIEDFFPLL	1448
RESULT 38			
ADJ68448			
ID	ADJ68448 standard; protein; 3225 AA.		
XX	ADJ68448;		
AC	AC		
XX	06-MAY-2004 (first entry)		
DT	Human heat mitochondrial protein as a therapeutic target SeqID254.		
DE	mitochondrial; human; screening assay; diabetes mellitus;		
XX	Huntington's disease; osteoarthritis;		
KW	Leber's hereditary optic neuropathy; LHON;		
KW	mitochondrial encephalopathy lactic acidosis and stroke; MELAS;		
KW	myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;		
KW	neuroprotective; nontropic; antidiabetic; anticonvulsant; antiarthritic;		
KW	osteopathic; ophthalmological; cytostatic.		
XX	Homo sapiens.		
OS	WO2003087768-A2.		
XX	23-OCT-2003.		
PD	04-APR-2003; 2003WO-USO10870.		
XX	12-APR-2002; 2002US-0372843P.		
PR	17-JUN-2002; 2002US-0389987P.		
PR	20-SEP-2002; 2002US-0412418P.		
XX	(MITO-) MITOKOR.		
PA	(BUCK-) BUCK INST AGE RES.		
XX	Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;		
PI	Warnock DB;		
XX	WPI; 2003-845369/78.		
DR	Identifying a mitochondrial target for drug screening assays and for		
XX	treating diseases associated with altered mitochondrial function,		
PT	comprises detecting a modified polypeptide in a sample and correlating		
PT	with the disease.		
XX	Claim 1; SEQ ID NO 254; 180pp; English.		
PS	This invention relates to novel mitochondrial targets that can be used		
XX	for therapeutic intervention in treating a disease associated with		
CC	altered mitochondrial function. Specifically, it refers to a method for		
CC	identifying proteins of the human heart mitochondrial proteome that are		
CC	useful for drug screening assays, as well as therapeutic targets. The		
CC	present invention describes a method for identifying such proteins that		
CC	can be used in the treatment of various diseases associated with altered		
CC	mitochondrial function including diabetes mellitus, Huntington's disease,		
CC	osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial		
CC	encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy		
CC	ragged red fibre syndrome (MERRF) or cancer. Accordingly, these		
CC	compositions have neuroprotective, nontropic, antidiabetic,		
CC	anticonvulsant, antiarthritic, osteopathic, ophthalmological and		
CC	cytostatic activities. This polypeptide sequence is a human heart		

CC	mitochondrial protein of the invention.		
XX	Query Match	2.8%; Score 271; DB 7; Length 3225;	
SQ	Best Local Similarity	18.8%; Pred No. 3.8e-06;	
	Matches 308; Conservative	246; Mismatches 507; Indels 578; Gaps 76;	
<hr/>			
QY	106	IPKYQVWISFEPEDKERFRGLFHLEKLEGNIAQSAATKFIYLLPLDMPKALGQYS	165
DB	1704	LSKFKQSLMSEKDSLSEVQ-----DLKHQIE-GNVSKQAN-----LEATEKHNDQTN	1750
QY	166	Y-----IVDKFNWLIITHPLSNFSAQSIKPLALTRSDPTAKLNQFNQDELWV	214
DB	1751	VTEEGTQSTIPGETEBOODLSMSRPTCTSESVPKAGNPAVSKD-----FSSHDEINN	1803
QY	215	YLEKFFDLEALKANIRLQTAADFSEFGKLVDPVYSPFIRNPQNKQEWASDLNODQKTVRL	274
DB	1804	YLOQ-----IDQKERI-----AGLEEEK-----QNKKEFSQTLNENKNTLLS	1841
QY	275	YLRTFSPQAKTILKDYKYKDETFLLSSIDLKASNGTSL-----PANENDLKQDLDVLDLV	330
DB	1842	QISTK-DGELK-MLQEVTKWNLNQIOBELSRVTKLKETAEEKDDLEERLMNQLAEL	1899
QY	331	SDYFGQSETITS-----NSQVPV-----PASERSLKORVKKKDQOKPRIEK	374
DB	1900	NGSIGNYQDVTDAQIKNELLSESEMNLKCVSELEEEKQQLVKBKTKVSEIRKEYLEK	1959
QY	375	F-SLYEYDALSFYSOLOELV-SKPNSIIDLVN-----ATLARNLR-----	412
DB	1960	IQQAQKEPGNKSHAKELQELKQEQEVKQLOKDCIRYQEKISALERTVKALEFVQTESQ	2019
QY	413	-----FSLGYNFLFDDLASHLDYFVLVSKAKIKQSSITKJFI	451
DB	2020	KOLEITKENLAQAVEHKKQAQELAFKVLDDTQSEA-ARVLADNLKLKELQSNKESV	2078
QY	452	ELPIK-----ISLKSSIIGDQPNIKTLFEKVTFKLNFNFR--DVEIEKAFGLLYPGVN---	503
DB	2079	KSQMKQKDELRLEQAEBKHLKE--KKMQEKLDAALREKVLHETTIGEIVTLNKKD	2136
QY	504	EELEQARK-----AQASPEK-----EKSKKGLAKERS-----QOKE	534
DB	2137	KEVQQLQENLDSTVTQLAATKMSLSQDDRDRVIDEAKKWKERFSDAIQSKEEIRLKE	2196
QY	535	ENSKAJNNQ-----EGLEEDDNITERLPENSPITQYQENAGLGASDPKPY	579
DB	2197	DNCSVLKQDLQWMSIHWEELKINISRLHDKQIWESKAO-TEVQLQKVCDTLQGENKE-	2254
QY	580	MIKDVQVQRYLAKSQIOELIKAKDYTKLAKLLSNRHTYNIURLKEQLFDVNPRIYPSR	639
DB	2255	LLSQLEBTRHLHYSSQ-----NELAKLESELKS-----LKDQLTDL-----SN	2292
QY	640	DIEKAKFVLDTKKNKYQIYSSASPVFNQKWSLFGVYRYLLGLDPKQTHLVLKLGKA	699
DB	2293	SLEKCK-----EOKGNLEGIIRQ-----QEA	2313
QY	700	GLQFE--GYENLPSDFNLE-----DLKNIRIKTFLFSQKDNFKLSLDDFNYYD	746
DB	2314	DIQNSKFSYEQLTDLQASRELTSLRHEEINMKEQKIISLSCKEERAIQVAIAELRQQHD	2373
QY	747	GEIKAPFGLPLPLPHELRRN-----SSNGGGSQNSPWEQEIISOFKQONLSNQDLAQ	802
DB	2374	KEIKE-----LENLSQSEEBENIVLEEBKKAVDKTNQ--LMETLTKIKKENIQQAQLDS	2427
QY	803	FSTKI-----WEKIGDENEFD-----ONNRL-----	824
DB	2428	FVKSMSSLQNRDRRIVGDYQQLQEEERHLSILEKDLQIQEAAENKLEIRGLRSHMDD	2487
QY	825	-----QYK-----LLKDLQESWI-----NKTRDNLWYTYLGDKLKVKP	857
DB	2488	LNSENAKDLAELLYREDLNQVITIKDSQKQLLEVLQQLQONKELENKY-AKLEEKLESE	2546
QY	858	KNN--LEAKF-----RQISNMQE-----	873

DB	2547	EANEDLRRSFNALQBEKQDLSKEISLKVSIISQLTQVTFALQEBEGTGLGHYHAQLKVKKEE	2606
QY	874	--LLTAFYTS-----AALSNMNYTQDSGAKSTIIFEEIAELDPKVKKEKVGADVYQLK-F	925
DB	2607	VHRLSALFSSSKQRIAELEELVCVQKEAAK-----KVGEIEDKLK-----ELKHL	2653
QY	926	HVAIFDDNAGKFNQEVIRSSRTIYLTGSGSKULEADTIDQLNQAVKNAPLGLQSF---	982
DB	2654	HHDAGIMRNETETABERVAELARDLV-----EMEQKLL-MVTKENKGLTAIQISFGRS	2705
QY	983	-----YLTERFGVFKLATSIAVQHKQKKTLPKKLANDGYTLIHDKLKKVIPQ	1033
DB	2706	MSLQNSRDHANEDELKRYKDYASLKAELAKBOGL-----LARE-----RDALLSETAFS	2757
QY	1034	ISSSEKOWFE-GKLNQNGSQSNVNVSTFGSIIESPY-----FS---TNFQEDAD-----	1079
DB	2758	MNSTEENSLSHLEKLNQQLLSKDEQLLLHSSQLEDSYNQVQSFASKAMASLQNERDHLWNE	2817
QY	1080	-----LDQDQDSDS-----ROGNNSLDNQEBAGLLKQKLAILLGNQFIQYQ	1120
DB	2818	LEKFRKSEBGKQSAAPSTSPAQVQSLKKAMSSLQNDRLKE-----LKNLQQQYLQ	2872
QY	1121	QN-----DKIEFEIINVE-----KVSLSF	1141
DB	2873	INOETELHPLKAQLOEQYODTKAFQIMQBELRQENLSWQHEDQLRMKSSWEIHERM	2932
QY	1142	RYEFKLAKTLENG-----KTIRVLSDETMSLIVNTTIETKTPMSAVPE-----VFD	1188
DB	2933	KEOYLMASIDKQOQLSHLQNLRELSSSQTPKQVQROQASPTSPDGSSQNLVYE	2992
QY	1189	TKWFOYDRTPLAAKTKFVLKPK--DOIYVDGSGNISDKWLASIPLVIHQOMLRLSPV	1245
DB	2993	TELL-----RTQLNDSLKEIHQELRIQOLNSNFSQLLEKNTLSIQLCDTSQSLR----	3043
QY	1246	VKTIRELGLKTEQQ-----QOQOQOQOQOOP-----OKKAV-RKEE	1281
DB	3044	-----ENQHYGDLNHCVALEKQVQELQAGPLNIDVAPGAPQEKNGVHRKSDP	3092
QY	1282	LETYNPKOBFNLNPLTKAHLRTLNLNVNDPNYKIEDUKVKNAGDHQL-----AFSLR	1337
DB	3093	EELRFPQQSFSQAQ-----QQLCNT--RQEVNELRKLLEERDQORVAENALSVA	3140
QY	1338	ANNIKRL-----MNTPI	1349
DB	3141	EQIIRRLHSEHSDSRTPI	3159
<hr/>			
RESULT 39			
ABU42656			
ID	ABU42656	standard; protein; 6641 AA.	
XX	AC	ABU42656;	
XX	DT	19-JUN-2003 (first entry)	
XX	DE	Protein encoded by Prokaryotic essential gene #28183.	
XX	KW	Antisense; prokaryotic essential gene; cell proliferation; drug design.	
XX	OS	Staphylococcus epidermidis.	
XX	PN	WO200277183-A2.	
XX	PD	03-OCT-2002.	
XX	PF	21-MAR-2002; 2002WO-US009107.	
XX	PR	21-MAR-2001; 2001US-00815242.	
XX	PR	06-SEP-2001; 2001US-00948993.	
XX	PR	25-OCT-2001; 2001US-0342923P.	
XX	PR	08-FEB-2002; 2002US-00072851.	
XX	PR	06-MAR-2002; 2002US-0362699P.	

This Page Blank (uspio)